

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 13:48:27 ; Search time 10429 Seconds
(without alignments)
11580.846 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaattccatttaaat.....tgagttccagctggccc 4150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.ph:*

8: gb.pl:*

9: gb.pr:*

10: gb.ro:*

11: gb.sts:*

12: gb.sy:*

13: gb.un:*

14: gb.vi:*

15: em.ba:*

16: em.fun:*

17: em.hum:*

18: em.in:*

19: em.mu:*

20: em.om:*

21: em.or:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vi:*

30: em.htg.hum:*

31: em.htg.inv:*

32: em.htg.other:*

33: em.htg.mus:*

34: em.htg.pin:*

35: em.htg.rod:*

36: em.htg.mam:*

37: em.htg.vit:*

38: em.sy:*

39: em.htg.hum:*

40: em.htg.mus:*

41: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4150	100.0	4150	9	AF320307	Homo sapi
2	4120.2	99.3	159351	2	AC018783	Homo sapi
3	4109.8	99.0	115278	2	AL139819	Human DNA
4	1992	48.0	185644	2	AL391723	Homo sapi
5	1932.8	46.6	1970	6	AX431388	Sequence
6	1448.2	34.9	179879	2	AC026883	Homo sapi
7	612.4	14.8	617	6	AX357255	Homo sapi
8	460	11.1	164201	9	AL157935	Sequence
9	447.2	10.8	183386	2	AC119725	Human DNA
10	433	10.4	192773	2	AL731547	Homo sapi
11	423.8	10.2	190508	9	AC005037	Homo sapi
12	415.2	10.0	44496	9	AC004760	Homo sapi
13	415.2	10.0	190821	9	AC006111	Homo sapi
14	408.8	9.9	183690	9	AC091180	Homo sapi
15	407.4	9.8	170528	9	AC083866	Homo sapi
16	407.4	9.8	217615	9	AC011005	Homo sapi
17	404.2	9.7	137289	9	AL161909	Human DNA
18	403	9.7	189179	2	AC090265	Homo sapi
19	402.8	9.7	181597	9	AC090515	Homo sapi
20	402.8	9.7	187734	9	AC025918	Homo sapi
21	394.2	9.5	179599	2	AC013278	Homo sapi
22	394.2	9.5	196894	2	AC007448	Homo sapi
23	392.2	9.5	155559	2	AL133336	Homo sapi
24	392.2	9.5	157581	2	AP005624	Homo sapi
25	391.6	9.4	171641	9	AC018512	Homo sapi
26	390	9.4	171555	2	AC087537	Homo sapi
27	390	9.4	181891	9	AC023356	Homo sapi
28	389.8	9.4	253178	2	AC025162	Homo sapi
29	388.4	9.4	172421	2	AC013279	Homo sapi
30	386.2	9.3	112364	9	AC109927	Homo sapi
31	386	9.3	151131	2	AC021945	Homo sapi
32	386	9.3	197092	2	AC036155	Homo sapi
33	385.8	9.3	147128	2	AC127512	Homo sapi
34	385.8	9.3	171936	2	AC120031	Homo sapi
35	385.8	9.3	191673	2	AL356743	Homo sapi
36	385.8	9.3	207418	9	AC090774	Homo sapi
37	385.8	9.3	208009	2	AC126365	Homo sapi
38	384.8	9.3	140529	2	AC011486	Homo sapi
39	384.2	9.3	167296	2	AC116346	Homo sapi
40	384	9.3	133475	9	AC006254	Homo sapi
41	382.4	9.2	188639	9	AC010997	Homo sapi
42	381.6	9.2	80775	9	AC068539	Homo sapi
43	381.2	9.2	150681	9	AC006011	Homo sapi
44	381.2	9.2	187064	9	AC011476	Homo sapi
45	381.2	9.2	204340	2	AC019238	Homo sapi

ALIGNMENTS

RESULT 1
AF320307
LOCUS AF320307
DEFINITION Homo sapiens stearyl-CoA desaturase (SCD) gene, exon 1 and intron 1 and partial cds.
ACCESSION AF320307
VERSION AF320307.1 GI:14150490
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4150)
AUTHORS Zhang L., Ge L., Tran, T., Stenn, K. and Prouty, S.M.
TITLE Isolation and characterization of the human stearyl-CoA desaturase

gene promoter: requirement of a conserved CCAAT cis-element
 Biochem. J. 357 (Pt 1), 183-193 (2001)
 MEDLINE 21308466
 PUBMED 11415448
 REFERENCE 2 (bases 1 to 4150)
 AUTHORS Zhang, L., Ge, L., Tran, T., Stenn, K. and Prouty, S.M.
 Direct Submission
 TITLE Submitted (09-NOV-2000) Skin Biology TRC, Johnson and Johnson,
 JOURNAL CPWW, 199 Grandview Road, Skillman, NJ 08558, USA
 FEATURES
 source
 1. 4150
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /tissue_type="scalp skin keratinocytes; liver; hair
 follicles; brain"
 3602..>4150
 /gene="SCD"
 3602..3606
 /gene="SCD"
 3869..3872
 /gene="SCD"
 3924..3927
 /gene="SCD"
 <3958..>4129
 /gene="SCD"
 /product="stearoyl-CoA desaturase"
 /note="3602"
 <3958..4129
 /gene="SCD"
 /number=1
 3958..4102
 /gene="SCD"
 4103..>4129
 /gene="SCD"
 /codon_start=1
 /product="stearoyl-CoA desaturase"
 /protein_id="AAK54510.1"
 /db_xref="GI:14150491"
 /translation="MPAHLQDD"
 4130..>4150
 /gene="SCD"
 /number=1
 BASE COUNT 988 a 1076 c 1065 g 1021 t
 ORIGIN
 Query Match 100.0%; Score 4150; DB 9; Length 4150;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGGAATTCATCCATTAAATCATACAAATTTAATGGCTTTTAGTATATTCACAGGTTGTGC 60
 DB 1 AGGAATTCATCCATTAAATCATACAAATTTAATGGCTTTTAGTATATTCACAGGTTGTGC 60
 QY 61 ATCCATCACATCCATTTTAGAACAGTTTATTACTCCAAAATAAACCCTGCATTCCTT 120
 DB 61 ATCCATCACATCCATTTTAGAACAGTTTATTACTCCAAAATAAACCCTGCATTCCTT 120
 QY 121 AGCCATCACCCCAACATCCATCCATCCCTCCAGCCCTGGGCAACCAACCAATCTAC 180
 DB 121 AGCCATCACCCCAACATCCATCCATCCCTCCAGCCCTGGGCAACCAACCAATCTAC 180
 QY 181 TTTCTGCTCTATAAATTGGCAATTCGGACATTCATATAAATGGNAGCAACACAT 240
 DB 181 TTTCTGCTCTATAAATTGGCAATTCGGACATTCATATAAATGGNAGCAACACAT 240
 QY 241 GTGAGACTTTGTGACTGGCTGCTTCACTTAGCATTCCTATTTTAAAGGCTCATTTATGTTA 300
 DB 241 GTGAGACTTTGTGACTGGCTGCTTCACTTAGCATTCCTATTTTAAAGGCTCATTTATGTTA 300
 QY 301 CAGTACTTAGCAGTACTTCAATCTTTTTTATCTCAAAATGGTATTCACACTGTGTGGGTAT 360
 DB 301 CAGTACTTAGCAGTACTTCAATCTTTTTTATCTCAAAATGGTATTCACACTGTGTGGGTAT 360

QY 361 CCATATCATATTATTAGAGACAGGTTCTCTACTGTCCACCAGGCTGGAGTGCAGTGGC 420
 DB 361 CCATATCATATTATTAGAGACAGGTTCTCTACTGTCCACCAGGCTGGAGTGCAGTGGC 420
 QY 421 ACAATCATAGCTCACTGTAACTCTAACTCTGGGCTCAAGTATCCCTACTACCTCAGCC 480
 DB 421 ACAATCATAGCTCACTGTAACTCTAACTCTGGGCTCAAGTATCCCTACTACCTCAGCC 480
 QY 481 TCCAGAGTAGCTAGGACTACAGGCACACACAGCCATACCTGGCTAAATTTTTTTTTTAAAT 540
 DB 481 TCCAGAGTAGCTAGGACTACAGGCACACACAGCCATACCTGGCTAAATTTTTTTTTTAAAT 540
 QY 541 TTTCAATTTATGTATTCATTTTCTTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 DB 541 TTTCAATTTATGTATTCATTTTCTTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
 QY 601 ACTTGTGTACCCAGGCTGGAGGCGAGTGGCATGGTGGAGCTGAGCAGCTGAGCAGCTTGCCT 660
 DB 601 ACTTGTGTACCCAGGCTGGAGGCGAGTGGCATGGTGGAGCTGAGCAGCTTGCCTTGCCT 660
 QY 661 GGGCTCAAGTAGTCTCTCTGCTCAGCCCTCCCAAGTAGCTGGAGCTACAAACACCTGTCA 720
 DB 661 GGGCTCAAGTAGTCTCTCTGCTCAGCCCTCCCAAGTAGCTGGAGCTACAAACACCTGTCA 720
 QY 721 CCATGCTGGCTGATATTTTCTTTTCTTTTGAACAGGGTATACACTCTGTGTGCCAGGCTGGA 780
 DB 721 CCATGCTGGCTGATATTTTCTTTTCTTTTGAACAGGGTATACACTCTGTGTGCCAGGCTGGA 780
 QY 781 GTACAGTGGCTGATAATAGCTCACTGAGCTCCCTCTCTGGGCTCAAGCAATCCGCTG 840
 DB 781 GTACAGTGGCTGATAATAGCTCACTGAGCTCCCTCTCTGGGCTCAAGCAATCCGCTG 840
 QY 841 GCCTCAGCATCCTGAGTAGCTGGAGCTACAGGCTTGTGCCACAGGCCAGCTAAGTTTT 900
 DB 841 GCCTCAGCATCCTGAGTAGCTGGAGCTACAGGCTTGTGCCACAGGCCAGCTAAGTTTT 900
 QY 901 AAAAAATGATTTTGGTATAGAGAGGCTTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGT 960
 DB 901 AAAAAATGATTTTGGTATAGAGAGGCTTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGT 960
 QY 961 TTGAGACAAGGCTCACTATGTGTCATGATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCA 1020
 DB 961 TTGAGACAAGGCTCACTATGTGTCATGATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCA 1020
 QY 1021 TCCTATCTGTTTCATTAGTACAGTGTGACAGCATTTAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
 DB 1021 TCCTATCTGTTTCATTAGTACAGTGTGACAGCATTTAGGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
 QY 1081 TGAATAATACCTCCAGTGAATTTTATGTATACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
 DB 1081 TGAATAATACCTCCAGTGAATTTTATGTATACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
 QY 1141 GTTGGGTTTATCTAGGAGTGGAAATGCTGGATCCCGGGTAAATTTTGTACAGGCGAG 1200
 DB 1141 GTTGGGTTTATCTAGGAGTGGAAATGCTGGATCCCGGGTAAATTTTGTACAGGCGAG 1200
 QY 1201 TTCAGGGGAAGAAACCTTGGGAAATGAAGCATGTTTGAAGATCAGCAAGAGTGCAGGG 1260
 DB 1201 TTCAGGGGAAGAAACCTTGGGAAATGAAGCATGTTTGAAGATCAGCAAGAGTGCAGGG 1260
 QY 1261 GTTTTTCGGAGTTTATTTTATTTTATTTCTGTGACAAATGTGCGAGTTTGTATGAAGATACAAG 1320
 DB 1261 GTTTTTCGGAGTTTATTTTATTTTATTTCTGTGACAAATGTGCGAGTTTGTATGAAGATACAAG 1320
 QY 1321 TTATACTAAGTGAAGTGAAGATTAAGGCTGGAAATAGGGCTTTCAGAGTAAAAATCATGA 1380
 DB 1321 TTATACTAAGTGAAGTGAAGATTAAGGCTGGAAATAGGGCTTTCAGAGTAAAAATCATGA 1380
 QY 1381 AGCACTTTGAATACCAAAATTAAGAGCTGTGGCTGTGAACAAAATTAATAAAAAATCAAA 1440
 DB 1381 AGCACTTTGAATACCAAAATTAAGAGCTGTGGCTGTGAACAAAATTAATAAAAAATCAAA 1440

QY	1441	TTTTTTTTTTTTTTTTTTGAGAAAGTCTTGGCTCTTTACCCCTGGCTGGAGGCGAGTGGTG	1501
Db	1441	TTTTTTTTTTTTTTTTTTGAGAAAGTCTTGGCTCTTTACCCCTGGCTGGAGGCGAGTGGTG	1500
QY	1501	TGATCTCAGCTCACTCAACTTTCCGCTCCCGGGTTCAAGCAATTCCTCTGCTCAGCCT	1560
Db	1501	TGATCTCAGCTCACTCAACTTTCCGCTCCCGGGTTCAAGCAATTCCTCTGCTCAGCCT	1560
QY	1561	CCCAAGTAGCTGGGACTACAGGCACCTCCCAACATGCCAGCTGATTTTTGTATTTTATG	1620
Db	1561	CCCAAGTAGCTGGGACTACAGGCACCTCCCAACATGCCAGCTGATTTTTGTATTTTATG	1620
QY	1621	TAGAGATGGGATTTCACCTTTGTTGGCCAAAGCTGGTCTCAAACTTTTTCGTGTCATAATTG	1680
Db	1621	TAGAGATGGGATTTCACCTTTGTTGGCCAAAGCTGGTCTCAAACTTTTTCGTGTCATAATTG	1680
QY	1681	TTCTAACTATTGTTCTCTTTGCTGAGGTAGGGCCCCCAGACCAAAAAAAATAAATCTTAG	1740
Db	1681	TTCTAACTATTGTTCTCTTTGCTGAGGTAGGGCCCCCAGACCAAAAAAAATAAATCTTAG	1740
QY	1741	AAATCCAAATCAGTGTGTGGTTTGACCACTGTCACTTGAGAACACACAGTGTGACCAAGGGC	1800
Db	1741	AAATCCAAATCAGTGTGTGGTTTGACCACTGTCACTTGAGAACACACAGTGTGACCAAGGGC	1800
QY	1801	CTCAGGAGTAGAGGTGATCTCTGCTCGAAAGAGAAAATAGAAATGAAATATTCCTCCGGGCC	1860
Db	1801	CTCAGGAGTAGAGGTGATCTCTGCTCGAAAGAGAAAATAGAAATGAAATATTCCTCCGGGCC	1860
QY	1861	AGCGTGTGGTCTATGCCTGTAAATCCACGACACTTGGGAGGCCAAGGCATGTGGATCAC	1920
Db	1861	AGCGTGTGGTCTATGCCTGTAAATCCACGACACTTGGGAGGCCAAGGCATGTGGATCAC	1920
QY	1921	CTCAGCTCAGGAGTTCAAAACCCAGCGTGGCCAAACATGATGTGAAACCCCGTCTCTACTAAAA	1980
Db	1921	CTCAGCTCAGGAGTTCAAAACCCAGCGTGGCCAAACATGATGTGAAACCCCGTCTCTACTAAAA	1980
QY	1981	ATACAAAAATTAGCTAAGTGTGGTGGCGCATGCCTGTAAATCCACGCTACTTGGGAGGGT	2040
Db	1981	ATACAAAAATTAGCTAAGTGTGGTGGCGCATGCCTGTAAATCCACGCTACTTGGGAGGGT	2040
QY	2041	GAGGCAGGAGAAATTTCTGAACCCGGGAGGACAGAGTTGCAGTCAACGGAGATCAACACCA	2100
Db	2041	GAGGCAGGAGAAATTTCTGAACCCGGGAGGACAGAGTTGCAGTCAACGGAGATCAACACCA	2100
QY	2101	CTGCATCTCAGCCTGGGGAGAGCGAGACTTCCTCTCAAAAAACAAAAACAAAAAGA	2160
Db	2101	CTGCATCTCAGCCTGGGGAGAGCGAGACTTCCTCTCAAAAAACAAAAACAAAAAGA	2160
QY	2161	ATTAAACAAATTAGACATTGCAGACAGACCTGAGGGGGTTCAGACCACCTACAGATTTC	2220
Db	2161	ATTAAACAAATTAGACATTGCAGACAGACCTGAGGGGGTTCAGACCACCTACAGATTTC	2220
QY	2221	TGTGCCACATGCCAAGTACTTCTGAGGCATGACTGGATGAGCTGTCCACATCTGAAATCA	2280
Db	2221	TGTGCCACATGCCAAGTACTTCTGAGGCATGACTGGATGAGCTGTCCACATCTGAAATCA	2280
QY	2281	TCCAGTCTGTTTCAGAACTTTCACACCCGACAGGAGCCAGGACTGGAATGCGAGTCTCCT	2340
Db	2281	TCCAGTCTGTTTCAGAACTTTCACACCCGACAGGAGCCAGGACTGGAATGCGAGTCTCCT	2340
QY	2341	GGTCACTGCCACAGAGTTGGCCTTGACCCCTGAGACCAGTGGCCCAACAAAGGAGCTGCTT	2400
Db	2341	GGTCACTGCCACAGAGTTGGCCTTGACCCCTGAGACCAGTGGCCCAACAAAGGAGCTGCTT	2400
QY	2401	AGTCTACTCTCCAGGAATGCCAGTGGTTGTCTCTCTGGGAAGTGAATCATTTGGCGCAG	2460
Db	2401	AGTCTACTCTCCAGGAATGCCAGTGGTTGTCTCTCTGGGAAGTGAATCATTTGGCGCAG	2460
QY	2461	CACCTCCGTAATTTCTCCTCTTCCCAGGGGAAGGATCTCTAGGGCAGTATTTGGGAAGACA	2520
Db	2461	CACCTCCGTAATTTCTCCTCTTCCCAGGGGAAGGATCTCTAGGGCAGTATTTGGGAAGACA	2520
QY	2521	TGGGCATGGAAGGACACCGGGTGAATGCATAGCCTGCTGGTCTGAGCTCTCATGCTAA	2580

[illegible]

Db 3601 GCCAATGGCAAGCGGACGAGGTGGACCAAAATTCCTTCGGCAATGACGAGCGGA 3660
 QY 3661 GTTTACAGACCTCATAGCATTTCCCGAGAGCAGGCGAGGCGAGGCGGGTGG 3720
 Db 3661 GTTTACAGACCTCATAGCATTTCCCGAGAGCAGGCGAGGCGAGGCGGGTGG 3720
 QY 3721 TGTGGTTCGGTTCGGCAGCATCCCGGGGCGCTGTGGTTCGGTTCGGCAGGCGCTCGGCC 3780
 Db 3721 TGTGGTTCGGTTCGGCAGCATCCCGGGGCGCTGTGGTTCGGTTCGGCAGGCGCTCGGCC 3780
 QY 3781 TCTGTCTCCCTCCCGCTCCCGGCTTACCTCCACGCGGAGCGCCCGGCGGCTCAACTCC 3840
 Db 3781 TCTGTCTCCCTCCCGCTCCCGGCTTACCTCCACGCGGAGCGCCCGGCGGCTCAACTCC 3840
 QY 3841 TCGCAGCTTGGCCCTGTTGGCAGCGATTAAGAGGGGCTGAGGAATACCGGACAGGT 3900
 Db 3841 TCGCAGCTTGGCCCTGTTGGCAGCGATTAAGAGGGGCTGAGGAATACCGGACAGGT 3900
 QY 3901 CACCCGTTGGCAGCTTAGCTTTAAATTCCTCGGCTCGGGAGCTCCACGACCGCGGCT 3960
 Db 3901 CACCCGTTGGCAGCTTAGCTTTAAATTCCTCGGCTCGGGAGCTCCACGACCGCGGCT 3960
 QY 3961 AGCGCCGACACAGCAGTGCAGCGCGCGGCTCAGCGGCTACCGCGGCGGCTTCG 4020
 Db 3961 AGCGCCGACACAGCAGTGCAGCGCGCGGCTCAGCGGCTACCGCGGCGGCTTCG 4020
 QY 4021 AAACCGAGTCCCTCCGCGGACCGGAACTCCGCTCCGGAGCTCAGCCCGCTCGGAAAGTG 4080
 Db 4021 AAACCGAGTCCCTCCGCGGACCGGAACTCCGCTCCGGAGCTCAGCCCGCTCGGAAAGTG 4080
 QY 4081 ATCCCGCATCCGAGACCGGAACTCCGCTCCGGAGCTCAGCCCGCTCGGAAAGTG 4140
 Db 4081 ATCCCGCATCCGAGACCGGAACTCCGCTCCGGAGCTCAGCCCGCTCGGAAAGTG 4140
 QY 4141 AGCCTGGGCC 4150
 Db 4141 AGCCTGGGCC 4150

RESULT 2
 AC018783/c
 LOCUS
 DEFINITION Homo sapiens clone RP11-2M13, WORKING DRAFT SEQUENCE, 12 unoriented
 pieces.
 AC018783
 VERSION AC018783.3 GI:7229991
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 159351)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 Melidrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,
 O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
 Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Subramanian,A., Takamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
 Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.
 and Zody,M.

TITLE
JOURNAL

COMMENT

Direct Submission
 Submitted (19-DEC-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 12, 2000 this sequence version replaced gi:6692332.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L2867
 Center clone name: 2.M.13
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; version 0.960731
 Assembly program: Phrap; version 0.960731
 Consensus quality: 15139 bases at least Q40
 Consensus quality: 155177 bases at least Q30
 Consensus quality: 157011 bases at least Q20
 Insert size: 167000; agarose-fp
 Insert size: 158251; sum-of-contigs
 Quality coverage: 4.9 in Q20 bases; agarose-fp
 Quality coverage: 5.2 in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently

consists of 12 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1. 2297: contig of 2297 bp in length
 2298 2397: gap of 100 bp
 2398 5719: contig of 3322 bp in length
 5720 5819: gap of 100 bp
 5820 10149: contig of 4330 bp in length
 10150 10249: gap of 100 bp
 10250 16496: contig of 6247 bp in length
 16497 16596: gap of 100 bp
 16597 21823: contig of 5227 bp in length
 21824 21923: gap of 100 bp
 21924 25747: contig of 3824 bp in length
 25748 25847: gap of 100 bp
 25848 31971: contig of 6124 bp in length
 31972 32071: gap of 100 bp
 32072 44449: contig of 12378 bp in length
 44450 44549: gap of 100 bp
 44550 64978: contig of 20429 bp in length
 64979 65078: gap of 100 bp
 65079 91491: contig of 26413 bp in length
 91492 91591: gap of 100 bp
 91592 121937: contig of 30346 bp in length
 121938 122037: gap of 100 bp
 122038 159351: contig of 37314 bp in length.

FEATURES

source

Location/Qualifiers
 1. .159351
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-2M13"
 /clone="RP11-2M13"
 1. .2297
 /note="assembly_fragment"
 2398. .5719
 /note="assembly_fragment"
 5820. .10149
 /note="assembly_fragment"
 10250. .16496
 /note="assembly_fragment"
 16597. .21823
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature

```
misc_feature /note="assembly_fragment"
21924..25747
/note="assembly_fragment"
clone_end:SP6
vector_side:right
25848..31971
/note="assembly_fragment"
clone_end:T7
vector_side:left
32072..44449
/note="assembly_fragment"
44550..64978
/note="assembly_fragment"
65079..91491
/note="assembly_fragment"
91592..121937
/note="assembly_fragment"
122038..159351
/note="assembly_fragment"
BASE COUNT 44283 a 34858 c 35376 g 43731 t 1103 others
ORIGIN

Query Match 99.3%; Score 4120.2; DB-2; Length 159351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4144; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 3 GAATTCATCCATTTAAATCATACAAATTAATGGCTTTTAGTATATTCACAGGTTGTGCAT 62
Db 59390 GAATTCATCCATTTAAATCATACAAATTAATGGCTTTTAGTATATTCACAGGTTGTGCAT 59331
Qy 63 CCATCACAAATCCATTTAGACAGTTTATTTACTCCAAAAATAAACCCCTGCATTCCTTAG 122
Db 59330 CCATCACAAATCCATTTAGACAGTTTATTTACTCCAAAAATAAACCCCTGCATTCCTTAG 59271
Qy 123 CCATCACCCCCCAACATCCTGCATCCTCCTCCAGCCCTGGGCAACCAACCAATCTATT 182
Db 59270 CCATCACCCCCCAACATCCTGCATCCTCCTCCAGCCCTGGGCAACCAACCAATCTATT 59211
Qy 183 TCTGCTCTATAAATTCGCCAATTCGACATTTTCATATAATGAAGCAACACACATGT 242
Db 59210 TCTGCTCTATAAATTCGCCAATTCGACATTTTCATATAATGAAGCAACACACATGT 59151
Qy 243 GAGACTTTGTGACGTGGCTTTCACITTAGCATTTCTATTTTAAAGCTCATTATGTTACA 302
Db 59150 GAGACTTTGTGACGTGGCTTTCACITTAGCATTTCTATTTTAAAGCTCATTATGTTACA 59091
Qy 303 GTACTTAGCAGTACTTCAATCTTTTATTCCTCAATGGTATTCACATGTTGGGTATCC 362
Db 59090 GTACTTAGCAGTACTTCAATCTTTTATTCCTCAATGGTATTCACATGTTGGGTATCC 59031
Qy 363 CATATCATATTATTAGACAGAGTTCTCACCTGTCCACCCAGGCTGGAGTGCAGTGGCAC 422
Db 59030 CATATCATATTATTAGACAGAGTTCTCACCTGTCCACCCAGGCTGGAGTGCAGTGGCAC 58971
Qy 423 AATCATAGCTCAGTGAACCTCAAACTCCTGGGCTCAAGTGATCCTACTACCTCAGCCTC 482
Db 58970 AATCATAGCTCAGTGAACCTCAAACTCCTGGGCTCAAGTGATCCTACTACCTCAGCCTC 58911
Qy 483 CAGAGTAGCTAGGACTACAGGCACACACAGCCATACCTGGCTAATTTTTTTTAAATTT 542
Db 58910 CAGAGTAGCTAGGACTACAGGCACACACAGCCATACCTGGCTAATTTTTTTTAAATTT 58851
Qy 543 TCATTTTATGATTCATTTCTTTCTTTTGTGTTGTTGTTGTTGATAGGCTCTCAC 602
Db 58850 TCATTTTATGATTCATTTCTTTCTTTTGTGTTGTTGTTGTTGATAGGCTCTCAC 58791
Qy 603 TTTGTTATCCAGGCTGGAGGCGAGTGGCATGGTGACAGCTGAGCAGCCTTGACTTCTGG 662
Db 58790 TTTGTTATCCAGGCTGGAGGCGAGTGGCATGGTGACAGCTGAGCAGCCTTGACTTCTGG 58731
Qy 663 GCTCAAGTAGTCTTCCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGGTGTCAAC 722
Db 58730 GCTCAAGTAGTCTTCCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGGTGTCAAC 58671
```

```
Qy 723 ATGCGCTGGCTGATATATTTTTTTTGAACAGGGTATCACTCTGTGTGCCAGGCTGGAGT 782
Db 58670 ATGCGCTGGCTGATATATTTTTTTTGAACAGGGTATCACTCTGTGTGCCAGGCTGGAGT 58611
Qy 783 ACAGTGGCGTAATAATAGTCACTGCAGCCTCCCTCCCTGGGCTCAAGCAATCCGCTGGC 842
Db 58610 ACAGTGGCGTAATAATAGTCACTGCAGCCTCCCTCCCTGGGCTCAAGCAATCCGCTGGC 58551
Qy 843 CTCAGCATCCTGAGTAGCTGGGACTACAGGCTGTGTCACCCAGGCGCCAGCTAAGTTTTAA 902
Db 58550 CTCAGCATCCTGAGTAGCTGGGACTACAGGCTGTGTCACCCAGGCGCCAGCTAAGTTTTAA 58491
Qy 903 AAAATGATTTTTGGTATAGAGAGGCTCTTGTCTATGTTGTCTAGGCTGTATTTTATTTGTT 962
Db 58490 AAAATGATTTTTGGTATAGAGAGGCTCTTGTCTATGTTGTCTAGGCTGTATTTTATTTGTT 58431
Qy 963 GAGACAAGGCTCACTATGTTGCCATGATCCCCCAACCTCCACTTCCCAAGGTGCTCATC 1022
Db 58430 GAGACAAGGCTCACTATGTTGCCATGATCCCCCAACCTCCACTTCCCAAGGTGCTCATC 58371
Qy 1023 TTATCTGTTTCATTAGTTCAGTTGACAGACATTTAGTGTGTTTCCACTTTTGTGACCATTAAG 1082
Db 58370 TTATCTGTTTCATTAGTTCAGTTGACAGACATTTAGTGTGTTTCCACTTTTGTGACCATTAAG 58311
Qy 1083 AATAATACCTCAGTGAATTAATTCATGATATACATTTGTGTGGGCAATGTTTTCATTTCTGT 1142
Db 58310 AATAATACCTCAGTGAATTAATTCATGATATACATTTGTGTGGGCAATGTTTTCATTTCTGT 58251
Qy 1143 TGGGTTTATATCTAGAGTGGAAATGCTGATCCCGGTAATATTTTGACAGGCGAGATT 1202
Db 58250 TGGGTTTATATCTAGAGTGGAAATGCTGATCCCGGTAATATTTTGACAGGCGAGATT 58191
Qy 1203 CAGGGGAAGAAAACTTGGGAAATGAAGCATGTTTAGAAATCAGCAAGAGTGCAGGGGT 1262
Db 58190 CAGGGGAAGAAAACTTGGGAAATGAAGCATGTTTAGAAATCAGCAAGAGTGCAGGGGT 58131
Qy 1263 TTTTCGGAGTTTATTTTATATTTCTGTGTGACAAATGTCAGTTTGTATGAAGATACAGTT 1322
Db 58130 TTTTCGGAGTTTATTTTATATTTCTGTGTGACAAATGTCAGTTTGTATGAAGATACAGTT 58071
Qy 1323 ATATCAAGTGAGAACTGAGAAATTAAGGCTGGAATTAAGGCTTCAGAGTAAAAATCATGAAG 1382
Db 58070 ATATCAAGTGAGAACTGAGAAATTAAGGCTGGAATTAAGGCTTCAGAGTAAAAATCATGAAG 58011
Qy 1383 CACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAAAATAATAAAAAATCAAAAT 1442
Db 58010 CACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAAAATAATAAAAAATCAAAAT 57951
Qy 1443 TTTTTCAGAGTTCAGAAAGAGTCTTGTCTTTTCCACTGCTGGGCGGAGTGGTGTG 1502
Db 57950 TTTTTCAGAGTTCAGAAAGAGTCTTGTCTTTTCCACTGCTGGGCGGAGTGGTGTG 57891
Qy 1503 ATCTCAGCTCACTGCAACTTTTCGCTCCCGGTTTCAAGCAATTCCTGCTTCAGGCTCC 1562
Db 57890 ATCTCAGCTCACTGCAACTTTTCGCTCCCGGTTTCAAGCAATTCCTGCTTCAGGCTCC 57831
Qy 1563 CAAGTAGCTGGGACTACAGGCACTTCCCAACATGCCAGTCCAGCTGATTTTGTATTTTATGTA 1622
Db 57830 CAAGTAGCTGGGACTACAGGCACTTCCCAACATGCCAGTCCAGCTGATTTTGTATTTTATGTA 57771
Qy 1623 GAGATGGGATTTCACTTTGTTGGGCAAGCTGGTCTCAAACTTTTGTCTGTATATTTGTT 1682
Db 57770 GAGATGGGATTTCACTTTGTTGGGCAAGCTGGTCTCAAACTTTTGTCTGTATATTTGTT 57711
Qy 1683 GTAACATTTGTTCCCTTTTGTCTGAGTGGGCGCCAGACACAAAAATAATACTTAGAA 1742
Db 57710 GTAACATTTGTTCCCTTTTGTCTGAGTGGGCGCCAGACACAAAAATAATACTTAGAA 57651
Qy 1743 TCCAAATCAGTGTGTGTTGTTGACCACTGTCTACCTTGAGAACCAAGTGTGACCAAGGCGCT 1802
Db 57650 TCCAAATCAGTGTGTGTTGTTGACCACTGTCTACCTTGAGAACCAAGTGTGACCAAGGCGCT 57591
```

QY 1803 CAGAGTAGAGGTGATCTCTGCTGAAAGAGAAATAGAAATGAAAAATATCTCCGGGCGAG 1862
Db 57590 CAGAGTAGAGGTGATCTCTGCTGAAAGAGAAATAGAAATGAAAAATATCTCCGGGCGAG 57591
QY 1863 GCGTGGTGGCTCATGCTCTGTAATCCACAGACTTTGGGAGGCGAAGGATGGAATCACT 1922
Db 57530 GCGTGGTGGCTCATGCTCTGTAATCCACAGACTTTGGGAGGCGAAGGATGGAATCACT 57471
QY 1923 GAGTTCAGGAGCTCAAAAACAGCTGGCCAAATGTTGAAACCCCGCTCTCTACTAAAAAT 1982
Db 57470 GAGTTCAGGAGCTCAAAAACAGCTGGCCAAATGTTGAAACCCCGCTCTCTACTAAAAAT 57411
QY 1983 AAAAAAATAGCTAAGTGTGGTGGCGCATGCCCTGTAATCCAGCTACTTTGGGAGGTGA 2042
Db 57410 AAAAAAATAGCTAAGTGTGGTGGCGCATGCCCTGTAATCCAGCTACTTTGGGAGGTGA 57351
QY 2043 GGCAGGAGAAATTTCTGAACCCGGGAGGACAGAGTTCAGTGAAGCGAGATCAACACACT 2102
Db 57350 GGCAGGAGAAATTTCTGAACCCGGGAGGACAGAGTTCAGTGAAGCGAGATCAACACACT 57291
QY 2103 GCATCCAGCCTGGGGAGAGAGAGACTTCCTCTCAAAAAAACAACAAAAAGAAAT 2162
Db 57290 GCATCCAGCCTGGGGAGAGAGAGACTTCCTCTCAAAAAAACAACAAAAAGAAAT 57231
QY 2163 TAACCAATTAGACATTCAGAGAGAACCTGAAGGGGTCAAGCAAGTACAGATTTCTG 2222
Db 57230 TAACCAATTAGACATTCAGAGAGAACCTGAAGGGGTCAAGCAAGTACAGATTTCTG 57171
QY 2223 TGCCACATGCCAAGTACTTTCAGAGCATGACTGGATGAGCTGCCATCATCTGAATCATC 2282
Db 57170 TGCCACATGCCAAGTACTTTCAGAGCATGACTGGATGAGCTGCCATCATCTGAATCATC 57111
QY 2283 CAGCTTGTTCAGAACTTTTCACACCGACAGAGGAGCTGGAATGCACTCTCTGG 2342
Db 57110 CAGCTTGTTCAGAACTTTTCACACCGACAGAGGAGCTGGAATGCACTCTCTGG 57051
QY 2343 TCATGCGCCAGAGAGTTGGCTTCACCTTGAGACCATGGCCAAAGAGAGCTGCTTAG 2402
Db 57050 TCATGCGCCAGAGAGTTGGCTTCACCTTGAGACCATGGCCAAAGAGAGCTGCTTAG 56992
QY 2403 TCTACCTCCAGGAAATCCAGGTGCTTGTCTTCTGGGAAGTGAATCATTTGGGCGAGCA 2462
Db 56991 TCTACCTCCAGGAAATCCAGGTGCTTGTCTTCTGGGAAGTGAATCATTTGGGCGAGCA 56932
QY 2463 CTCGTAATTTCTCTCTCCAGGGGAAGATCCTAGGCGAGTATTTGGGAAAGACATG 2522
Db 56931 CTCGTAATTTCTCTCTCCAGGGGAAGATCCTAGGCGAGTATTTGGGAAAGACATG 56872
QY 2523 GGCATGAAGGACACCGGCTGAATGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2582
Db 56871 GGCATGAAGGACACCGGCTGAATGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56812
QY 2583 CTCCTACAGACACGGAAGATGGGGCACAGGACAGATCAGTAGGCTCAGAGCATCTC 2642
Db 56811 CTCCTACAGACACGGAAGATGGGGCACAGGACAGATCAGTAGGCTCAGAGCATCTC 56752
QY 2643 AGGACCGAGGGCAATATGCTGCTGACAGGATTAAGAGCTTGGGCTCTCATATGCTGT 2702
Db 56751 AGGACCGAGGGCAATATGCTGCTGACAGGATTAAGAGCTTGGGCTCTCATATGCTGT 56692
QY 2703 TTTGGGCTCAACTGCGAGCTCCGCTCACTTACTGTTGCTGTGACCATGGCAAGTTATT 2762
Db 56691 TTTGGGCTCAACTGCGAGCTCCATCACTTACTGTTGCTGTGACCATGGCAAGTTATT 56632
QY 2763 CCATCTCTCATATCTCTTCTCTCACTTTTAAATGGAATTAATGGGTACCCACTCCCA 2822
Db 56631 CCATCTCTCATATCTCTTCTCTCACTTTTAAATGGAATTAATGGGTACCCACTCCCA 56572
QY 2823 GGGTCACAGAGAGCTTACAGAAAAGATTTCTGTGAATTTGGCTTGCAGTAATTAATCAA 2882
Db 56571 GGGTCACAGAGAGCTTACAGAAAAGATTTCTGTGAATTTGGCTTGCAGTAATTAATCAA 56512
QY 2883 TACCTGCCAGCTATTCTTATTCCACATCCAAAGCCCTTTCGCTGCTGCTGGGTGAAACAA 2942

Db 56511 TACCTGCCAGCTATTCTTATTCCATCAAGCCCTTTTCGCTGCTGCTGGGTGAAACAA 56452
QY 2943 CATGTCACTGTTTCTGAGGCTTTCCACAAAGAAATTTCCAAAAATTAACAACTGCCAGTC 3002
Db 56451 CATGTCACTGTTTCTGAGGCTTTCCACAAAGAAATTTCCAAAAATTAACAACTGCCAGTC 56392
QY 3003 TGAAGAAATCTCAAAAACATCCCGACAGCATCTTGAGGGGGGGCTTGGGATGAGGACTG 3062
Db 56391 TGAAGAAATCTCAAAAACATCCCGACAGCATCTTGAGGGGGGGCTTGGGATGAGGACTG 56332
QY 3063 CCGGCCCGGGTCTCTGAACAGGATGCTGCGCGAGGACACACACAGCAGCAGCTCTG 3122
Db 56331 CCGGCCCGGGTCTCTGAACAGGATGCTGCGCGAGGACACACACAGCAGCAGCTCTG 56272
QY 3123 TGTGGGCGCGAGTCCCGTGGCTGCCGGTGAAGCAGCGCTGCTGCTGGGGGGGCGAG 3182
Db 56271 TGTGGGCGCGAGTCCCGTGGCTGCCGGTGAAGCAGCGCTGCTGCTGGGGGGGCGAG 56212
QY 3183 AGCCATTGTTGCGAGGCGTACCGAGCCCCCGCGCTCGCCGGGAGGAGGCGGGGCTTC 3242
Db 56211 AGCCATTGTTGCGAGGCGTACCGAGCCCCCGCGCTCGCCGGGAGGAGGCGGGGCTTC 56152
QY 3243 CCGGCT-CCCCAAGCTCCAGATCCTGGGGTGGCTGCCACAGTCTCCTGCCACGGGCTTG 3301
Db 56151 CCGGCTCCCCAAGCTCCAGATCCTGGGGTGGCTGCCACAGTCTCCTGCCACGGGCTTG 56092
QY 3302 GGGACGGGAAGACGGGACGAGATGTTAGTGGTGGGGGGCCCCCGAGGGTTACCACTG 3361
Db 56091 GGGACGGGAAGACGGGACGAGATGTTAGTGGTGGGGGGCCCCCGAGGGTTACCACTG 56032
QY 3362 TTTCTCTGAGAAATTTCCCGAGTGCACCCACCGTCTTCTCGTGTGCCCGAGGCGCGT 3421
Db 56031 TTTCTCTGAGAAATTTCCCGAGTGCACCCACCGTCTTCTCGTGTGCCCGAGGCGCGT 55972
QY 3422 CTGGGCTAGGCTCCGCGCCCCAGCCCCAAACCCGGTCCCCAGCCCCCTTCCAGAGAGAA 3481
Db 55971 CTGGGCTAGGCTCCGCGCCCCAGCCCCAAACCCGGTCCCCAGCCCCCTTCCAGAGAGAA 55912
QY 3482 CTCCCGAGCGGGATCCGGGACAGGCGCCAGCGGGGTGGAAGAGAGCTGAGAGGA 3541
Db 55911 CTCCCGAGCGGGATCCGGGACAGGCGCCAGCGGGGTGGAAGAGAGCTGAGAGGA 55852
QY 3542 GAAACAGAGGGGAGGGGAGGAGAGTGGCGGACAGAGGGAACAGAGATTTGCCCGGAG 3601
Db 55851 GAAACAGAGGGGAGGGGAGGAGAGTGGCGGACAGAGGGAACAGAGATTTGCCCGGAG 55792
QY 3602 CCAATGCAACGGGACAGAGTGGACCAATTCCTTCGGGCAATGAGAGCGCGAG 3661
Db 55791 CCAATGCAACGGGACAGAGTGGACCAATTCCTTCGGGCAATGAGAGCGCGAG 55732
QY 3662 TTTACAGAAAGCTCATTTAGCATTTCCCGAGAGGCGGGGAGGGGAGGCGGGTGGT 3731
Db 55731 TTTACAGAAAGCTCATTTAGCATTTCCCGAGAGGCGGGGAGGGGAGGCGGGTGGT 55672
QY 3722 GTGGTGGTGGTGGGAGCATCCCCGGGCGCTGCTGCGGTGCGCGAGGCGCTCGGCT 3781
Db 55671 GTGGTGGTGGTGGGAGCATCCCCGGGCGCTGCTGCGGTGCGCGAGGCGCTCGGCT 55612
QY 3782 CTGCTCTCTCCCTCCCGCTTACCTCCAGGGGAGCGCGCGCGCGCAGTCACTCT 3841
Db 55611 CTGCTCTCTCCCTCCCGCTTACCTCCAGGGGAGCGCGCGCGCGCAGTCACTCT 55552
QY 3842 CGCACTTTGGCCCTGCTTGGCAGGGATTAAGGGGCTGAGGAATACCGGACAGGTC 3901
Db 55551 CGCACTTTGGCCCTGCTTGGCAGGGATTAAGGGGCTGAGGAATACCGGACAGGTC 55492
QY 3902 ACCGTTGCGAGCTAGCTTTAAATTTCCCGCTCGGGGACCTCCACGACCGCGGCTA 3961
Db 55491 ACCGTTGCGAGCTAGCTTTAAATTTCCCGCTCGGGGACCTCCACGACCGCGGCTA 55432
QY 3962 GCGCCACACAGCTAGCTGAGCGGCGCGGCTCAGCGGTACCGCGGCTTCGA 4021

Db 55431 GCGCGCAACACAGTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGGGGGCTTCCA 55372
 QY 4022 AACCGAGTCTCCGGGACCCGAACTCCGCTCGGAGCCTCAGCCCGCTGGAAAGTGA 4081
 Db 55371 AACCGAGTCTCCGGGACCCGAACTCCGCTCGGAGCCTCAGCCCGCTGGAAAGTGA 55312
 QY 4082 TCCCGGATCCGAGACCAAGATCCCGCCACATCTGCTGCAGGAGGATGAGTTTCCCA 4141
 Db 55311 TCCCGGATCCGAGACCAAGATCCCGCCACATCTGCTGCAGGAGGATGAGTTTCCCA 55252
 QY 4142 GCCTGGGCC 4150
 Db 55251 GCCTGGGCC 55243

RESULT 3
 AL139819
 LOCUS
 DEFINITION Human DNA sequence from clone Rpl1-34D15 on chromosome 10, complete sequence.
 ACCESSION AL139819
 VERSION HTG.
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE Tracey, A.
 AUTHORS Direct Submission
 TITLE Submitted (18-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 JOURNAL requests: clonerequest@sanger.ac.uk
 COMMENT On Sep 19, 2001 this sequence version replaced gi:14586033.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; Sw.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
 Rpl1-34D15 is from the library RPl1-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone Rpl1-34D15 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true right end of clone Rpl1-34D15 is at 115278 in this sequence. The true left end of clone Rpl1-285F16 is at 56886 in this sequence. The true right end of clone Rpl1-316M21 is at 2000 in this sequence.

FEATURES
 source
 Location/Qualifiers
 1..115278
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="Rpl1-34D15"

BASE COUNT 31458 a 26875 c 25750 g 31195 t
 ORIGIN
 Query Match 99.0%; Score 4109.8; DB 9; Length 115278;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 4144; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
 QY 3 GAATTCATCCCAATTTAAATCATACAAATTTAAATGGCTTTTAGTATATTCACAGGTTGTGCAT 62
 Db 64592 GAATTCATCCATTTAAATCATACAAATTTAAATGGCTTTTAGTATATTCACAGGTTGTGCAT 64651
 QY 63 CCATCACAATCCATTTTAAAGACAGTTTATTACTCCAAAATAAAACCCCTGCATTCCTTAG 122
 Db 64652 CCATCACAATCCATTTTAAAGACAGTTTATTACTCCAAAATAAAACCCCTGCATTCCTTAG 64711
 QY 123 CCATCACCACCCCAACATCTCCATCTCCTTCCAGCCCTGGGCAACCAACCAATCTACTT 182
 Db 64712 CCATCACCACCCCAACATCTCCATCTCCTTCCAGCCCTGGGCAACCAACCAATCTACTT 64771
 QY 183 TCTGTCTCTATAAAATTTGCCAAATTTCTGGACATTTTCATATAATGAAGCAACAACATGT 242
 Db 64772 TCTGTCTCTATAAAATTTGCCAAATTTCTGGACATTTTCATATAATGAAGCAACAACATGT 64831
 QY 243 GAGACTTTGTGACTGGCTGCTTTCACCTTAGCATTTCTATTTTAAAGCTCATTAATGTTACA 302
 Db 64832 GAGACTTTGTGACTGGCTGCTTTCACCTTAGCATTTCTATTTTAAAGCTCATTAATGTTACA 64891
 QY 303 GTACTTAGCAGTACTTCAATCTTTTATCTCAATGATGATCCACGTGTGGGTATCC 362
 Db 64892 GTACTTAGCAGTACTTCAATCTTTTATCTCAATGATGATCCACGTGTGGGTATCC 64951
 QY 363 CATATCATATTATTAGAGACAGTTTCTCAGCTCTGTCAACCCAGGCTGGAGTGCAGTGGCAC 422
 Db 64952 CATATCATATTATTAGAGACAGTTTCTCAGCTCTGTCAACCCAGGCTGGAGTGCAGTGGCAC 65011
 QY 423 AATCATAGCTCAGTGTAACTCAAACTCCTGGCTCAAGTGAATCTACTACCTCAGCCTC 482
 Db 65012 AATCATAGCTCAGTGTAACTCAAACTCCTGGCTCAAGTGAATCTACTACCTCAGCCTC 65071
 QY 483 CAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGCTGAGTGTGAGTGTGAGTGTGAGTGT 542
 Db 65072 CAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGCTGAGTGTGAGTGTGAGTGTGAGTGT 65131
 QY 543 TCATTTTATGATTCATTTCTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 602
 Db 65132 TCATTTTATGATTCATTTCTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 65191
 QY 603 TTTGTTACCCAGGCTGGAGGCGAGTGGCATGGTGACAGCTGAGCAGCCTTGACCTTCCTGG 662
 Db 65192 TTTGTTACCCAGGCTGGAGGCGAGTGGCATGGTGACAGCTGAGCAGCCTTGACCTTCCTGG 65251
 QY 663 GCTCAAGTGTATCTCTCCGCTCAGCTCCCAAGTAGCTGGGACTACAAACAGCTGTGACC 722
 Db 65252 GCTCAAGTGTATCTCTCCGCTCAGCTCCCAAGTAGCTGGGACTACAAACAGCTGTGACC 65311
 QY 723 ATGCTGCTGCTGATATTTTCTTTTGTGAACAGGATGATCCTCTGTGTTGCCAGGCTGGAGT 782
 Db 65312 ATGCTGCTGCTGATATTTTCTTTTGTGAACAGGATGATCCTCTGTGTTGCCAGGCTGGAGT 65371
 QY 783 ACAGTGGCGTAAATATAGCTACCTGACGCTCCCTCTCTGGGCTCAAGCAATCCGCTGGC 842
 Db 65372 ACAGTGGCGTAAATATAGCTACCTGACGCTCCCTCTCTGGGCTCAAGCAATCCGCTGGC 65431
 QY 843 CTCAGCATCTGAGTAGCTGGGACTACAGGCTTGTGCCACCCAGGCTGAGCTTAAGTTTAA 902
 Db 65432 CTCAGCATCTGAGTAGCTGGGACTACAGGCTTGTGCCACCCAGGCTGAGCTTAAGTTTAA 65491
 QY 903 AAAATGATTTTGTGATAGAGGAGTCTTGCTATGTTGCTCAGGCTGATTTTTATTGTT 962
 Db 65492 AAAATGATTTTGTGATAGAGGAGTCTTGCTATGTTGCTCAGGCTGATTTTTATTGTT 65551
 QY 963 GAGACAGGCTCTCACTATGTTGCCATGATCCCCCAGCTCCCAAGTCTCATC 1022

QY 3183 AGCAATGTTCCAGCGGTACGAGCCCGCCCGCTCCCGCGGAGGAGGCGGGCTTC 3242
|||||
Db 67771 AGCAATGTTCCAGCGGTACGAGCCCGCCCGCTCCCGCGGAGGAGGCGGGCTTC 67830
|||||
QY 3243 CGCGCT-CCCCAAGCTCCAGATCTTGGGTGGCTGCTCCAGCTCTCCTGCGACGCGCTGG 3301
|||||
Db 67831 CGCGCTCCCCAAGCTCCAGATCTTGGGTGGCTGCTCCAGCTCTCCTGCGACGCGCTGG 67890
|||||
QY 3302 GGGGACGGGAAGACGGGAGCGAGATGTTAGTGGTGGGCGCCCGCGAGGTTTACCACTG 3361
|||||
Db 67891 GGGGAC-GGAAGACGGGAGCGAGATGTTAGTGGTGGGCGCCCGCGAGGTTTACCACTG 67949
|||||
QY 3362 TTTCCTGAGAACTTCCCGAGTCCCGACCCACCCCGCTTCTCGTGGCGGAGGCGGCTG 3421
|||||
Db 67950 TTTCCTGAGAACTTCCCGAGTCCCGACCCACCCCGCTTCTCGTGGCGGAGGCGGCTG 68009
|||||
QY 3422 CTGGCTAGGCTCCGCGCCCGACCGCCCAACCGGGTCCCGACCGCTTCCAGAGAGAAAG 3481
|||||
Db 68010 CTGGCTAGGCTCCGCGCCCGACCGCCCAACCGGGTCCCGACCGCTTCCAGAGAGAAAG 68069
|||||
QY 3482 CTCCGAGCGGGGATCCGCGGCGAGAGCGCCAGCGGGGTGGAAGAGAGCTGAGAAGGA 3541
|||||
Db 68070 CTCCGAGCGGGGATCCGCGGCGAGAGCGCCAGCGGGGTGGAAGAGAGCTGAGAAGGA 68129
|||||
QY 3542 GAAACAGAGGGGAGGGGAGCGAGAGCTGGCGGAGAGGGAAACAGCAGATTGCGCCGAG 3601
|||||
Db 68130 GAAACAGAGGGGAGGGGAGCGAGAGCTGGCGGAGAGGGAAACAGCAGATTGCGCCGAG 68189
|||||
QY 3602 CCAATGCAACGGCAGGAGCGAGGTGGCCACCAATTTCCCTTCCGGCCCAATGACGAGCGGAG 3661
|||||
Db 68190 CCAATGCAACGGCAGGAGCGAGGTGGCCACCAATTTCCCTTCCGGCCCAATGACGAGCGGAG 68249
|||||
QY 3662 TTTCAGAGCGCTCATTTAGCATTTCCCGAGGAGCGGCGGAGGCGGAGCGCGGCTGT 3721
|||||
Db 68250 TTTCAGAGCGCTCATTTAGCATTTCCCGAGGAGCGGCGGAGGCGGAGCGCGGCTGT 68309
|||||
QY 3722 GTGGTGTGGGTGCGGAGCATCCCGCGCCCTGCTGCGGTGCGCGGAGCGCTCGGCT 3781
|||||
Db 68310 GTGGTGTGGGTGCGGAGCATCCCGCGCCCTGCTGCGGTGCGCGGAGCGCTCGGCT 68369
|||||
QY 3782 GTGTCTCTCCCGCTCCCGCGCTTACCTCCAGCGGGGACCGCGCGCCGCGCACTCCT 3841
|||||
Db 68370 GTGTCTCTCCCGCTCCCGCGCTTACCTCCAGCGGGGACCGCGCGCCGCGCACTCCT 68429
|||||
QY 3842 CGCACTTGGCGCTGCTGGCAGCGATTAAGAGGGGCTGAGGAATATCCGAGACAGGTC 3901
|||||
Db 68430 CGCACTTGGCGCTGCTGGCAGCGGATAAAGAGGGGCTGAGGAATATCCGAGACAGGTC 68489
|||||
QY 3902 ACCGTTGCCAGCTAGCTTAAATTTCCCGGCTCGGGGACCTCCAGCACCAGCGGCTA 3961
|||||
Db 68490 ACCGTTGCCAGCTAGCTTAAATTTCCCGGCTCGGGGACCTCCAGCACCAGCGGCTA 68549
|||||
QY 3962 GCGCGACACAGCTAGCTGTCAGAGCGCGCGGCTCAGCGCGTACCGGGGCTTCCA 4021
|||||
Db 68550 GCGCGACACAGCTAGCTGTCAGAGCGCGCGGCTCAGCGCGTACCGGGGCTTCCA 68609
|||||
QY 4022 AACCGAGCTCTCCGCGACCGCGAGCTCCGCTCGGAGCTCAGCCCGCTGGAAGTGA 4081
|||||
Db 68610 AACCGAGCTCTCCGCGACCGCGAGCTCCGCTCGGAGCTCAGCCCGCTGGAAGTGA 68669
|||||
QY 4082 TCCCGGATCCGAGACCAAGATCGCGCCGCTGCTGCGAGGAGCTGAGTTTCCA 4141
|||||
Db 68670 TCCCGGATCCGAGACCAAGATCGCGCCGCTGCTGCGAGGAGCTGAGTTTCCA 68729
|||||
QY 4142 GCTTGGCCC 4150
|||||
Db 68730 GCTTGGCCC 68738
|||||

RESULT 4
AL391723/c
LOCUS AL391723 185644 bp DNA linear HTG 11-SEP-2001

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Homo sapiens chromosome 10 clone RP11-3B21, *** SEQUENCING IN
PROGRESS ***, 14 unordered pieces.
AL391723
AL391723.7 GI:10186824
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Barton, J.
Direct Submission
Submitted (10-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 19, 2000 this sequence version replaced gi:10178830.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA38B21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 175002 bases at least Q40
Consensus quality: 179431 bases at least Q30
Consensus quality: 182156 bases at least Q20
Insert size: 184344; sum-of-contigs
Insert size: 140788; 6.0 error; agarose-fp
Quality coverage: 4.15x in Q20 bases; sum-of-contigs Quality
coverage: 6.51x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* I 59110: contig of 59110 bp in length
* 59111 59210: gap of 100 bp
* 59211 61285: contig of 2075 bp in length
* 61286 61385: gap of 100 bp
* 61386 64379: contig of 2994 bp in length
* 64380 64479: gap of 100 bp
* 64480 66747: contig of 2268 bp in length
* 66748 66847: gap of 100 bp
* 66848 114355: contig of 47508 bp in length
* 114356 114455: gap of 100 bp
* 114456 125406: contig of 10951 bp in length
* 125407 125506: gap of 100 bp
* 125507 127673: contig of 2167 bp in length
* 127674 127773: gap of 100 bp
* 127774 130111: contig of 2338 bp in length
* 130112 130211: gap of 100 bp
* 130212 132354: contig of 2143 bp in length
* 132355 132454: gap of 100 bp
* 132455 136537: contig of 4083 bp in length
* 136538 136637: gap of 100 bp
* 136638 139127: contig of 2490 bp in length
* 139128 139227: gap of 100 bp
* 139228 181355: contig of 42128 bp in length
* 181356 181455: gap of 100 bp
* 181456 183524: contig of 2069 bp in length
* 183525 183624: gap of 100 bp
* 183625 185644: contig of 2020 bp in length.
Location/Qualifiers
1. 185644
/organism="Homo sapiens"

[illegible]


```
QY 3467 CTTCCAGAGAGAAAGCTCCGACGCGGGATGCGGGGCGAGGCCGAGCGCGGGTGGAG 3526
Db 182146 CTTCCAGAGAGAAAGCTCCGACGCGGGATGCGGGGCGAGGCCGAGCGCGGGTGGAG 182087
QY 3527 AGAAGCTGAGAAGAGAAACAGAGAGGGAGGGGAGCGAGGAGCTGGCGGCGAGAGGAACA 3586
Db 182086 AGAAGCTGAGAAGAGAAACAGAGAGGGAGGGGAGCGAGGAGCTGGCGGCGAGAGGAACA 182027
QY 3587 GCAGATGCGCGGAGCAATGSCAACGCGAGCAGGAGTGGACCAAAATTCCTTCGCGC 3646
Db 182026 GCAGATGCGCGGAGCAATGSCAACGCGAGCAGGAGTGGACCAAAATTCCTTCGCGC 181967
QY 3647 AATGACGAGCGGAGTTTACAGAGCCATTAGCATTTCCCGAGGAGCGGAGGG 3706
Db 181966 AATGACGAGCGGAGTTTACAGAGCCATTAGCATTTCCCGAGGAGCGGAGGG 181907
QY 3707 CAGAGCCGCGGTGTTGTTGTCGCGAGCATCCCGCGCCCTTCGTCGCGTCC 3766
Db 181906 CAGAGCCGCGGTGTTGTTGTCGCGAGCATCCCGCGCCCTTCGTCGCGTCC 181847
QY 3767 CCGAGCCTCGGCTGCTGTCCTCCCGCTCCCGGCTTACCTCCAGCGGGAGCCCGC 3826
Db 181846 CCGAGCCTCGGCTGCTGTCCTCCCGCTCCCGGCTTACCTCCAGCGGGAGCCCGC 181787
QY 3827 CGCAGTCAACTCCTCGACCTTTCCTGCTGCGAGCATCCCGCGCCCTTACCTCCAGCGGGAGCTCAGGAA 3886
Db 181786 CGCAGTCAACTCCTCGACCTTTCCTGCTGCGAGCATCCCGCGCCCTTACCTCCAGCGGGAGCTCAGGAA 181727
QY 3887 ATACCGGACAGGTCACCGCTTTCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTC 3946
Db 181726 ATACCGGACAGGTCACCGCTTTCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTC 181667
QY 3947 CACGACCGCGGTAGCGCGGACACAGCTAGCGTGAAGCGCGCGGCTCAGCGGT 4006
Db 181666 CACGACCGCGGTAGCGCGGACACAGCTAGCGTGAAGCGCGCGGCTCAGCGGT 181607
QY 4007 ACCGCGGCGCTTCGAAACCGAGTCTCCGCGGACCCGAACTCCGCTCCGAGCCICAG 4066
Db 181606 ACCGCGGCGCTTCGAAACCGAGTCTCCGCGGACCCGAACTCCGCTCCGAGCCICAG 181547
QY 4067 CCGCTGAAAGTATCCCGGATCCGAGAGCCAGATGCCGGCCACTTGTGCGAGAC 4126
Db 181546 CCGCTGAAAGTATCCCGGATCCGAGAGCCAGATGCCGGCCACTTGTGCGAGAC 181487
QY 4127 GATGTGAGTTTCCAGCGCTGCGC 4150
Db 181486 GATGTGAGTTTCCAGCGCTGCGC 181463

RESULT 5
AX431388
LOCUS
DEFINITION Sequence 21 from Patent WO0240666.
ACCESSION AX431388
VERSION AX431388.1 GI:21656235
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1.
AUTHORS Enkins,D.K., Winther,M.D., Haardt,M., Goldberg,Y.P., Nwaka,S.O.,
Ponton,A., Allen,S.J., de Antueno,R.J. and Knickie,L.C.
TITLE Fat regulated genes, uses thereof, and compounds for modulating
same
JOURNAL Patent: WO 0240666-A 21 23-MAY-2002;
XENON GENETICS INC (CA)
FEATURES
Location/Qualifiers
1..1970
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 428 a 590 c 588 g 364 t
ORIGIN
```

```
Query Match 46.6%; Score 1932.8; DB 6; Length 1970;
Best Local Similarity 99.7%; Pred No. 0;
Matches 1967; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY 2121 GAGAGCGAGACTTCTCTCAAAAAACAAAAAAGAAATTAAGCAATTAAGCAATTG 2180
Db 1 GAGAGCGAGACTTCTCTCAAAAAACAAAAAAGAAATTAAGCAATTAAGCAATTG 60
QY 2181 CAGAGAGAACCTGAGAGGGGTGAGACAGTACAGATTTCTGTGCCACATGCCAAGTACT 2240
Db 61 CAGAGAGAACCTGAGAGGGGTGAGACAGTACAGATTTCTGTGCCACATGCCAAGTACT 120
QY 2241 TCTGAGGCACTGACTGAGTGTGTCACATCTGAATCATCCAGTCTTGTTCAGAACTT 2300
Db 121 TCTGAGGCACTGACTGAGTGTGTCACATCTGAATCATCCAGTCTTGTTCAGAACTT 180
QY 2301 TCACACCGGACAGGAGGAGCAGTGTGAATGAGTCTCTGTGCTACTGCCAGAGAGTTG 2360
Db 181 TCACACCGGACAGGAGGAGCAGTGTGAATGAGTCTCTGTGCTACTGCCAGAGAGTTG 240
QY 2361 GCCTTGACCTGAGACACAGTGGCCAAACAAAGAGCTGTAGTCTACCTCCAGGAAATC 2420
Db 241 GCCTTGACCTGAGACACAGTGGCCAAACAAAGAGCTGTAGTCTACCTCCAGGAAATC 299
QY 2421 CCAGGTGCTTCTCTCTGGAAGTGAATCAATTGGCGAGCAGTCCGTATTTCTCTCT 2480
Db 300 CCAGGTGCTTCTCTCTGGAAGTGAATCAATTGGCGAGCAGTCCGTATTTCTCTCT 359
QY 2481 TCCAGGGAAGATCTTAGGCGAGTATTTGGAAAGACATGGGCATGGAAGACACCGG 2540
Db 360 TCCAGGGAAGATCTTAGGCGAGTATTTGGAAAGACATGGGCATGGAAGACACCGG 419
QY 2541 GTGAATGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2600
Db 420 GTGAATGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
QY 2601 AGATGGGCGACAGGACAGATCAGTAGGCTCAGAGCATCTCAGGACCGAGGCAATAT 2660
Db 480 AGATGGGCGACAGGACAGATCAGTAGGCTCAGAGCATCTCAGGACCGAGGCAATAT 539
QY 2661 GGTCTCTGAGCAGGATTAAGAGCTTGGGCTCTCATATGCTGCTGCTGCTGCTGCTGCT 2720
Db 540 GGTCTCTGAGCAGGATTAAGAGCTTGGGCTCTCATATGCTGCTGCTGCTGCTGCTGCT 599
QY 2721 GCTCCGTCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2780
Db 600 GCTCCATCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
QY 2781 TTTCTCACTTTTAAATGGAATTAATGGGTACCCACCTCCAGGCTCAGAGAGGCTTA 2840
Db 660 TTTCTCACTTTTAAATGGAATTAATGGGTACCCACCTCCAGGCTCAGAGAGGCTTA 719
QY 2841 CAGAAACGATTTCTGTAATGGCTTGCAGTAATAATCAATACCTGCGACGCTATCTT 2900
Db 720 CAGAAACGATTTCTGTAATGGCTTGCAGTAATAATCAATACCTGCGACGCTATCTT 779
QY 2901 ATTCCACATCCAAGCCCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2960
Db 780 ATTCCACATCCAAGCCCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839
QY 2961 CGTTTCCCAAAAGAGATTTCCAAATTAACACCTGCGAGCTGGAAGATTTCCAAACA 3020
Db 840 CGTTTCCCAAAAGAGATTTCCAAATTAACACCTGCGAGCTGGAAGATTTCCAAACA 899
QY 3021 TCCGCGACGCTCCTGGAGGCGCGGCTTGGGATGGGACTGCCCGCCGCTCTGAAAC 3080
Db 900 TCCGCGACGCTCCTGGAGGCGCGGCTTGGGATGGGACTGCCCGCCGCTCTGAAAC 959
QY 3081 AGGATGCGTGGCGGCGGACACACACACACACACACACACACACACACACACACAC 3140
Db 960 AGGATGCGTGGCGGCGGACACACACACACACACACACACACACACACACACACAC 1019
```

QY	3141	TCCGGTCCGGGTGAGCAGCGCGTGGCTGGTGGGGGGGCGAGAGCCATTGTTTCGAGGCG	3200
Db	1020	TCCGGTCCGGGTGAGCAGCGCGTGGCTGGTGGGGGGGCGAGAGCCATTGTTTCGAGGCG	1079
QY	3201	TACGAGCCCCCGCGCTCGCCGGAGAGGAGCGGGCTTCCCGCT-CCCCAAGCTCC	3259
Db	1080	TACGAGTCCCGCGCTCGCCGGAGAGGAGCGGGCTTCCCGCTCCCGCAAGCTCC	1139
QY	3260	AGATCCTGGGTGGTGCACGCTTCCCTGCCACCGCTCGGGGAGCGGGAAGACGGGA	3319
Db	1140	AGATCCTGGGTGGTGCACGCTTCCCTGCCACCGCTCGGGGAGC-GGGAAGACGGGA	1198
QY	3320	CGAGATGTTAGTGGTGGGCGCCCCCGAGGGTTACCACTGTTTCTTGAGAACTTCCC	3379
Db	1199	CGAGATGTTAGTGGTGGGCGCCCCCGAGGGTTACCACTGTTTCTTGAGAACTTCCC	1258
QY	3380	CAGTGCACACCCCGTTCTCCGTGTGCCCGAGGGCCGGTCTTGGGCTAGCTCCGGCG	3439
Db	1259	CAGTGCACACCCCGTTCTCCGTGTGCCCGAGGGCCGGTCTTGGGCTAGCTCCGGCG	1318
QY	3440	CCCAGCCCCAACCGGGTCCCGACGCCCTTCCAGAGAGAAAGCTCCCGACGCGGATGCC	3499
Db	1319	CCCAGCCCCAACCGGGTCCCGACGCCCTTCCAGAGAGAAAGCTCCCGACGCGGATGCC	1378
QY	3500	GGGCGAGAGCCCCAGCGGGGTGGAAGAGAAGCTGAGAAGGAGAAACAGAGGGGAGGGG	3559
Db	1379	GGGCGAGAGCCCCAGCGGGGTGGAAGAGAAGCTGAGAAGGAGAAACAGAGGGGAGGGG	1438
QY	3560	AGCGAGAGCTGGCGGAGAGGGACACAGATTTCGCCCGAGACCAATGGCAACGGCAGGA	3619
Db	1439	AGCGAGAGCTGGCGGAGAGGGACACAGATTTCGCCCGAGACCAATGGCAACGGCAGGA	1498
QY	3620	CGAGGTGGCCACCAATTCCCTTCGGCCAAACACGAGCGCGAGTTTACAGAAGCCTCATT	3679
Db	1499	CGAGGTGGCCACCAATTCCCTTCGGCCAAACACGAGCGCGAGTTTACAGAAGCCTCATT	1558
QY	3680	GCATTTCCCGAGAGGCGAGGGCGAGAGCGGGTGGTGGTGGTGGTGGTGGTGGTGGCA	3739
Db	1559	GCATTTCCCGAGAGGCGAGGGCGAGAGCGGGTGGTGGTGGTGGTGGTGGTGGTGGCA	1618
QY	3740	GCATCCCGGGCCCTGCTGGGGTGGCGGAGAGCTTGGCCCTCTGTCTCTCCCGCTCCC	3799
Db	1619	GCATCCCGGGCCCTGCTGGGGTGGCGGAGAGCTTGGCCCTCTGTCTCTCTCCCGCTCCC	1678
QY	3800	GCCCTTACTCTCACCGCGGACCGCGCGCGAGTCAACTCTCTCGACCTTTGCCCTCGTT	3859
Db	1679	GCCCTTACTCTCACCGCGGACCGCGCGCGAGTCAACTCTCTCGACCTTTGCCCTCGTT	1738
QY	3860	GGCAGCGGATAAAGGGGGGTGAGAAATACCGGACACGGTACCCGTTGCCAGCTCTAG	3919
Db	1739	GGCAGCGGATAAAGGGGGGTGAGAAATACCGGACACGGTACCCGTTGCCAGCTCTAG	1798
QY	3920	CCTTTAAATTCGGGCTCGGGGACCTCCAGCGACCGCGGTAGCGCGACCAACAGCTAG	3979
Db	1799	CCTTTAAATTCGGGCTCGGGGACCTCCAGCGACCGCGGTAGCGCGACCAACAGCTAG	1858
QY	3980	CGTGCAGGCGCGCGGCTCAGCGCGTACCGCGGGGCTTCGAAACCGCAGTCTCCGGCG	4039
Db	1859	CGTGCAGGCGCGCGGCTCAGCGCGTACCGCGGGGCTTCGAAACCGCAGTCTCCGGCG	1918
QY	4040	ACCCCGAATTCGGTCCGGAGCCTCAGCCCCCTGGAAAGTATCCCCGCATC 4091	
Db	1919	ACCCCGAATTCGGTCCGGAGCCTCAGCCCCCTGGAAAGTATCCCCGCATC 1970	
RESULT 6			
AC026883			
LOCUS			
DEFINITION			
Homo sapiens chromosome 3 clone RP11-799L19 map 3p, WORKING DRAFT			
SEQUENCE, 41 unordered pieces.			
AC026883			
ACCESSION			
AC026883.3 GI:8101285			
KEYWORDS			
HTG; HTGS_PHASE1; HTGS_DRAFT.			

```
* 12254 14386: contig of 2133 bp in length
* 14387 14486: gap of unknown length
* 14487 16990: contig of 2504 bp in length
* 16991 17090: gap of unknown length
* 17091 19022: contig of 1932 bp in length
* 19023 19122: gap of unknown length
* 19123 20497: contig of 1375 bp in length
* 20498 20597: gap of unknown length
* 20598 22209: contig of 1612 bp in length
* 22210 22310: gap of unknown length
* 22310 24677: gap of unknown length
* 24678 24777: gap of unknown length
* 24778 27656: contig of 2878 bp in length
* 27657 27756: gap of unknown length
* 27757 30189: contig of 2434 bp in length
* 30190 30289: gap of unknown length
* 30290 33658: contig of 3369 bp in length
* 33659 33758: gap of unknown length
* 33759 36912: contig of 3154 bp in length
* 36913 37011: gap of unknown length
* 37012 41069: contig of 4057 bp in length
* 41070 41169: gap of unknown length
* 41170 43658: contig of 2489 bp in length
* 43659 46574: contig of 2816 bp in length
* 46575 46674: gap of unknown length
* 46675 49890: contig of 3216 bp in length
* 49891 49990: gap of unknown length
* 49991 53751: contig of 3761 bp in length
* 53752 53851: gap of unknown length
* 53852 57378: contig of 3527 bp in length
* 57379 57478: gap of unknown length
* 57479 61598: contig of 4120 bp in length
* 61599 61698: gap of unknown length
* 61699 65686: contig of 3988 bp in length
* 65687 65786: gap of unknown length
* 65787 70673: contig of 4887 bp in length
* 70674 70773: gap of unknown length
* 70774 76443: contig of 5669 bp in length
* 76443 76542: gap of unknown length
* 76543 80193: contig of 3651 bp in length
* 80194 80293: gap of unknown length
* 80294 83630: contig of 3337 bp in length
* 83631 83731: gap of unknown length
* 83732 89197: contig of 5466 bp in length
* 89198 89296: gap of unknown length
* 89297 93491: contig of 4195 bp in length
* 93492 93591: gap of unknown length
* 93592 99952: contig of 6361 bp in length
* 99953 100052: gap of unknown length
* 100053 107237: contig of 7185 bp in length
* 107238 107337: gap of unknown length
* 107338 113752: contig of 6415 bp in length
* 113753 113852: gap of unknown length
* 113853 119355: contig of 5503 bp in length
* 119356 119455: gap of unknown length
* 119456 127315: contig of 7860 bp in length
* 127316 127415: gap of unknown length
* 127416 137358: contig of 9943 bp in length
* 137359 137458: gap of unknown length
* 137459 148311: contig of 10853 bp in length
* 148312 148411: gap of unknown length
* 148412 160062: contig of 11651 bp in length
* 160063 160162: gap of unknown length
* 160163 179879: contig of 19717 bp in length.
FEATURES
  source
    1. .179879
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="3"
      /map="3p"
      /clone="RP11-799119"
      1. .1496
misc_feature
```

```
misc_feature /note="assembly_name:Contig7"
1597. .2949 /note="assembly_name:Contig9"
3050. 4409 /note="assembly_name:Contig10"
4510. .6137 /note="assembly_name:Contig11"
6238. 7284 /note="assembly_name:Contig12"
clone_end.SP6
vector_side:right"
7385. .9013 /note="assembly_name:Contig13"
9114. 10444 /note="assembly_name:Contig14"
10545. 12153 /note="assembly_name:Contig15"
12254. 14386 /note="assembly_name:Contig16"
14487. 16990 /note="assembly_name:Contig17"
17091. 19022 /note="assembly_name:Contig18"
19123. 20497 /note="assembly_name:Contig19"
20598. 22209 /note="assembly_name:Contig20"
22310. 24677 /note="assembly_name:Contig21"
24778. 27655 /note="assembly_name:Contig22"
27756. 30189 /note="assembly_name:Contig23"
30290. 33658 /note="assembly_name:Contig24"
33759. 36912 /note="assembly_name:Contig25"
37013. 41069 /note="assembly_name:Contig26"
41170. 43658 /note="assembly_name:Contig27"
43759. 46574 /note="assembly_name:Contig28"
46675. 49890 /note="assembly_name:Contig29"
49991. 53751 /note="assembly_name:Contig30"
53852. 57378
```

Query Match 34.9%; Score 1448.2; DB 2; Length 179879;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1456; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
QY 3 GAATTCATCCATTAAATCATACAAATTAATGGCTTTTAGTATATATACAGTGTGTCAT 62
|||||
Db 109955 GAATTCATCCATTAAATCATACAAATTAATGGCTTTTAGTATATATACAGTGTGTCAT 110014
QY 53 CCATCACAAATCCATTTAGAAGAGTTTATTACTCCAAAATAAACCCTGCATTCCTTAG 122
|||||
Db 110015 CCATCACAAATCCATTTAGAAGAGTTTATTACTCCAAAATAAACCCTGCATTCCTTAG 110074
QY 123 CCATCACCCGCCAACATCCCTCCATCCCTCTTCCAAAGCCCTGGGCAACCAATCTACTT 182
|||||
Db 110075 CCATCACCCGCCAACATCCCTCCATCCCTCTTCCAAAGCCCTGGGCAACCAATCTACTT 110134
QY 133 TCTGTCTCTATAAATTTGCCAATTTCCGACATTTTCATATATTAATGAAGCAACACATGT 242
|||||
Db 110135 TCTGTCTCTATAAATTTGCCAATTTCCGACATTTTCATATATTAATGAAGCAACACATGT 110194
QY 243 GAGACTTTGTGACGTGGCTGCTTTCACATTTAGCATTTCTATTTTAAAGGCTCATATTATGTACA 302
|||||
Db 110195 GAGACTTTGTGACGTGGCTGCTTTCACATTTAGCATTTCTATTTTAAAGGCTCATATTATGTACA 110254
```

[illegible]

D	b	111335	CACCTTGAATACCAAAATTAAAGAGCTTGCGCTGTATAACAAAATTAATAAAAAATCAACAATT	111394
Q	y	1443	TTTTITTTTTTTTTTGGAGAAGAGTCTTGTC	1471
D	b	111395	TTTTTTTAAATAAACGTCATAGTCCAGC	111423
R	E	S	U	L
RESULT 7				
LOCUS	AX357255		617 bp	DNA
DEFINITION	Sequence 1 from Patent WO0162954.			
ACCESSION	AX357255			
VERSION	AX357255.1	GI:18674409		
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Brownlie,A.J., Hayden,M.R., Attie,A.D., Ntambi,J.M., Gray-Keller,M.P. and Miyazaki,M.			
TITLE	Methods and compositions using stearyl-coa desaturase to identify triglyceride reducing therapeutic agents			
JOURNAL	Patent: WO 0162954-A 1 30-AUG-2001; Xenon Genetics Inc. (CA) ; WISCONSIN ALUMNI RESEARCH FOUNDATION (US) ; University of British Columbia (CA)			
FEATURES	Location/Qualifiers			
source	1..617			
BASE COUNT	117 a 211 c 205 g	84 t		
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606"			
Query Match	14.8%; Score 612.4; DB 6; Length 617;			
Best Local Similarity	99.8%; Pred.No.1.5e-139;			
Matches 613; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			
Q	y	3459	CCCAGCCCCCTCCACAGAGAAAGCTCCCGACGCCGATCCCGGCAGAGCCCGAGCGCG 3518	
D	b	4	CCCCGCCCTCCACAGAGAAAGCTCCCGACCGGGATCCCGGCAGAGCCCGAGCGCG 63	
Q	y	3529	GGTGAAGAGAAAGCTGAGAAGGAGAAACAGAGGGAGGGGAGCGAGAGCTGGCGGCAG 3578	
D	b	64	GGTGAAGAGAAAGCTGAGAAGGAGAAACAGAGGGAGGGGAGCGAGAGCTGGCGGCAG 123	
Q	y	3579	AGGGAAACAGCATTTGCGCCGAGCCAATGGCAACGCAGCAGAGTGCACCAAATCC 3638	
D	b	124	AGGGAAACAGCATTTGCGCCGAGCCAATGGCAACGCAGCAGAGTGCACCAAATCC 183	
Q	y	3639	CTTCGGCCAAATGACAGCGCGGAGTTTACAGAAAGCCCTATTAGCATTTCCCCAGAGCAG 3698	
D	b	184	CTTCGGCCAAATGACAGCGCGGAGTTTACAGAAAGCCCTATTAGCATTTCCCCAGAGCAG 243	
Q	y	3699	GGCAGGGCACAGCGCGGT 3758	
D	b	244	GGCAGGGCACAGCGCGGT 303	
Q	y	3759	GGGGTCTCGCGGAGCGCTCGGCCCTCTGTCTCTCCGCCCTCCCGCCCTTACTCTCAGCGCG 3818	
D	b	304	GGGGTCTCGCGGAGCGCTCGGCCCTCTGTCTCTCCGCCCTCCCGCCCTTACTCTCAGCGCG 363	
Q	y	3819	ACC GCCCGCGCGAGTCAACTCTTCGCATTTGCGCCCTCTGTTGGCAGCGGATAAAAAGGGG 3878	
D	b	364	ACC GCCCGCGCGAGTCAACTCTTCGCATTTGCGCCCTCTGTTGGCAGCGGATAAAAAGGGG 423	
Q	y	3879	CTGAGGAATACCGGACAGCGTFCACCCGCTGCCACCTAGCCTTAATTCGCCGCTCG 3938	
D	b	424	CTGAGGAATACCGGACAGCGTFCACCCGCTGCCACCTAGCCTTAATTCGCCGCTCG 483	
Q	y	3939	GGGAGCTCCAGGCACCGCGGCTAGCGCGGACAAACAGCTAGCGTGCAGAGCGCGCGGT 3998	
D	b	484	GGGAGCTCCAGGCACCGCGGCTAGCGCGGACAAACAGCTAGCGTGCAGAGCGCGCGGT 543	

Db	5965	TTTTTTAGGCTCGT--CAAGTGAAGCAATGACAGTGGAGAGGAACAAAGCAATCTGTAA	6022
QY	1371	AAATCATGAGCAGCTTTGAATACCAAAATTAAGAGAGCTGGCTGTAAACAAATATAAA	1430
Db	6023	CTAGTTGTGATCAATTACTGTGTAACACAAATGTTTTTTTCATTTAGTGTGCCATTAG	6082
QY	1431	AAAATCACAAATTTTTTTTTTTTTTTTTTTTGGAGAGAGACTGTGCTCTTCACCTGCTGGAG	1490
Db	6083	TTTTTTTTCATTAATTAATTAATTTTTTTTGGAGTGGAGTCTGTGCTGTGT--TGCCAGCGCTGGAA	6141
QY	1491	GGCAGTGGTGTGATCTCAGCTCAGCTGCAACTTTGCGCTCCCGGGTTCAAGCAATTTCTCCT	1550
Db	6142	TGCAATGGCGCGATCTCGSCTCACTGCAACCTCCACCTCTCTGGGTTCCAGTGATCTCCT	6201
QY	1551	GCTTCAGCCTCCCAAGTAGCTGGGACTACAGGCATTCACGCACTTCCCAACCATGCCAGCTGATTTT	1610
Db	6202	GCCTTAGCCTCCCAAGTAGCTGGGACTACAGGTGCATGCCACCATGCCAGCTAAATTTT	6261
QY	1611	GTATTTTTAGTAGAGATGGATTTACATTTGTTGGCAAGCTGTCFCAACT-----	1663
Db	6262	GTATTTTTAGTAAGATGGGTTTCGCCATGTTGCCAGGATGGGCTCAAACTCCTTGACG	6321
QY	1664	-----TTTTGCTGTCAATAATTTGTTAACATAATTTGTTCTTTTGTGAGTAGGGCCGCCAG	1719
Db	6322	TCAGGTGATCCACCGCCTTGGCCCTCCCAAAGTGCCTAGGATTAATAGCGTGAAGCCACTGC	6381
QY	1720	ACCAAAAAAATAATCTTAGAATCCAAATCAGTGTGTG-----GTTTGACCA	1768
Db	6382	ACCGCGCGCTCTCACATTTTTCACACATTTTGTGTAGCACATCCTCCCTAGCACAT	6441
QY	1769	CTGTCACTTGAACACACAGTGTGTACCAGGCGCTCAGAGTGTAGAGTGATCTCTGCTCGA	1828
Db	6442	TTTTTCTAACCATTATCCCAAGAGAAAAGAACACAGCAATCTACAAGCAGTGTCTATT	6501
QY	1829	AGAGAAATPAGAAATGAAAATATPCT-----	1853
Db	6502	GAAGAATGACCTCAAGACATTTCTGAACTAGTGGCATGTGAACAGTGAAGTCTCGAGAC	6561
QY	1854	-----CCGGGCGCAGCGGTGGTGGCTCATGCTGTAAATCCAGACACTTTGGGAGGCCA	1905
Db	6562	TAAACAGAGTAGACACGCGTGGTGGCTGTAGCGCTGTAAATCCACAGCACTTTGGGAGGCTG	6621
QY	1906	AGSCATGTGGATCACTCAGGTCTAGGAGTTCAAAACACAGCCTGCCAACATGGTGAACCC	1965
Db	6622	AGGTGAGCGGATCACTCAGGTTGGGAGTTCAAGACCAACCTGACCAACGTTGGAGAAACC	6681
QY	1966	CGTCTCTACTAAAAATACAAAAATAGCTAAGTGTGGTGGCGCATGCGTGTAAATCCCA	2025
Db	6682	CGGTCTCTACTAAAAATACTA--AAITAGCCAGGTGTGGTGGCGCATGCGTGTAAATCCCA	6739
QY	2026	GCTACTTGGGAGGTGAGCGAGGAGATTTCTTGA-----CCGGGAGGACGAGGTTCAGTGT	2084
Db	6740	GCTACTCGGGAGGCTGAGGCGAGGAGATCGCTTGAAGCCAGGAGCGGAGTGTGTGTG	6799
QY	2085	AAGCGAGATCACACACACTGCACTCCAGCCTGGG--GGAGAGAGCGAGACTTCCTCTCAAAA	2143
Db	6800	AGCCAAGATCGGCATTGACCTCTAGCCTGGGACGACAGAGTGAACCTCCGCGCTCAAAA	6859
QY	2144	AAACAAAAACAAAAGAA	2161
Db	6860	AAAAAAAAAAAAAAAAAAAA	6877

RESULT 9
AC119725 LOCUS
DEFINITION
AC119725 AC119725 183386 bp DNA linear HTG 09-MAY-2002
Homo sapiens chromosome 3 clone RP11-759D18, *** SEQUENCING IN
PROGRESS ***, 24 unordered pieces.
AC119725
AC119725.1 GI:20376806
HTG: HTGS_PHASE1.
Homo sapiens.
SOURCE

ОБЪЯВЛЕНИЕ

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 183386)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimige,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Briveva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinsh,H.J., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.H., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lounsged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,B., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 183386)

Worley,K.C.

Direct Submission

Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 183386)

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HGIR
 Center clone name: RP11-759D18
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 169013 bases at least Q40
 Consensus quality: 174569 bases at least Q30
 Consensus quality: 178019 bases at least Q20

Df 163996 ACCAAGCCTAGCTAAATTTTGTGTAATTTTACTAGAGATGGGGTTTCTCCATGTTGGCCAG 164005

QY 776 CTGAGTACAGTGGCGCTAAATAAGCTCACTGCAGCGCTCCCTCTGGGCTCAAGCAATC 835

Df 164056 CTGGCTCGA-----ACTCCTGACCTCAGGTGGTTC 164085

QY 836 CGCTGGCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTTGTGCC--ACAGGCCCGAGCT 893

Df 164086 CACCACCTGACCTCCACAGTCTGGGAATACAGGCTTGAGCCCGGCGACACCTGGCT 164145

QY 894 AAGTTTAAAAAATGATTTTGGTATAGAGAGAGGTCTTGCTATGTTGCTCAGGCTGATTT 953

Df 164146 AATTTTTTAAAAAATTTTTFAGAGA--CGGCTCTGCTTGTCAACCCAGGTGGAGT 164204

QY 954 TTTATTTGTTGAGACA-----AGGTCCTCACTATGTTGCCATGATCCCGCCACCTCCAC 1005

Df 164205 TCAGTTATGCCATCATAGCTCACTGTAAGTTCGAATTCGAGTCTGGGCTAAAGCAATCTCTCT 164264

QY 1006 TTCCCAAGTGCTCATCTTATCTGTTTCATTAGTCAGTTGACACACATTTAGTGTGTTTCC 1065

Df 164265 GCGTCACCTTCCCAAAATGCTGGGATACAGGCATGAGCCATGACCTGGCCAGTTTG 164324

QY 1066 ACITTTTGGACCATTAUGAATAAATACTCCAGTGAATATCTCATGTAATACATTTGTTGGGCA 1125

Df 164325 ACTTCTAGTCCACTTCTCGACAGCTCCCGAGCTCAAGTAAGTTCATTTCTCRACTCT 164384

QY 1126 TATGTTTTTCATTTCTGTTGGTT--TATATCTAGGAGTGGAAATGCTGTGGATCCGGGTAT 1184

Df 164385 TTGCTCTTGGACAAGCAAACTCTTCTGCATGGGTCCTCAGTCCCTTCTTACCCCTAT 164444

QY 1185 ATTTTGCAGGCAGAGTTCAAGGGGAAGAAAACCTGGGAAATGAAGCATGTTTAGAAT 1244

Df 164445 GTTCAGGCAAAACAGATTAATAATTAATACTAACATCCTCAGTGTATATCCAGGATCT 164504

QY 1245 CAGCAAGAGTCGAGGGTTTTTCGGAGTTTATTTTATATCTGTTGACAATGTCAGT 1304

Df 164505 CAGATCTTGCCTTAACTCTCATCTCTCTCTATATACATATATATATATACCCCT 164564

QY 1305 TTGATGAAGATACAAGTTTACTTAAGTGAGAAGTGAGATTAAGGCTGGAATAGGGCGTT 1364

Df 164565 GACTCCCTCCTCCGCTTTCACATTTTTCCTATTGATACATATATTTTACATGTTT 164624

QY 1365 CAGAGTAAATCATGAGCACTTTGAATACCAAAATTAAGGACT-----TGGCTGTAA 1418

Df 164625 ATGGGGTACATGTAATTTGCTACATGATAGAATATGATAGAGGTGGAACAATCCATCAC 164684

QY 1419 ACAAATAAATAAAAAATCAAAATTTTTTTTTTTTTTTTTCAGAAAGACTCTGCTCTTCA 1478

Df 164685 CCATTACCTTGAGTATTATCTTTTTTTTTTTTTTTTGGATGGAGTCTGCTGTGTG 164744

QY 1479 CCCTGGCTGAGGGCAGTGGTGTGATCTCAGTCACTGCACTTTCGCTCCCGGGTTCA 1538

Df 164745 CTAGGCTGGAGTGCAATGGCAAAATCTCAGTTTCACTGCAACCTCTGCTCCCGAGTTCA 164804

QY 1539 AGCAATTCCTCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAGGCCTTCCACCAATGCC 1598

Df 164805 AGTCATTCCTCGCTCAGCCCTCCTCAGTAGCTGGGATTAAGGCGTGTGCCACCAACCC 164864

QY 1599 CAGCTGAATTTGTATTTTTFAGTAGAGTGGGATTTTCACTTTGTTGGCAAGCTGCTCTC 1658

Df 164865 CGGCTAATTTTGTATTTTTFAGTAGAGCGGGCTTCACTATGTTGGTCAGGCTGGTCTC 164924

QY 1659 AAATTTTTTCTGCTCATATATGTTGTAACTATTTGTTCCCTTTTGTCTG---AGTAGGGCC 1714

Df 164925 AAATCTTAACCTCAGCGGATCCACCCCTCAGCCTCCCAAAAGTCTACAATATATAGGC 164984

QY 1715 CCCAGACCAAAAAATAAATCTTAGAATCCAAATCAGTGTGTGGTTTGTGACCACTGTCA 1774

Df 164985 GTGAGCCACTGCACCTGGCCAAATATTTATCATTTCTATGTTGTGGAATATTTCAAGTC 165044

QY 1775 CTTGAGAACCAAGTGTGACCGAGGGCTCAGGAGTAGAGGTATC-----TCTGCTCGAA 1830

Df 165045 CTCCTTCTATCTGCTTTGAAATATATCTACACATTTCTGCTGCTATAGCTGGCTCCACT 165104

Qy	1831	GAGAAATGAAATGAAATATTC - TCCGGGCCAGGCGTGGTGGCTCATGCTGTAAATCCCA	1889
Db	165105	ATCAACATCAATAATTTTCTTTCTTAGCCAGGCGTGGTGGCTCACGCTGTAAATCCCA	165164
Qy	1890	GCACTTTCGGGAGGCCAAGGCATGTGGATCACTGAGGTCAGGAGTTCAAAACACGACCTGG	1949
Db	165165	GCACTTTCGGGAGGCCAAGCAGGCAGATCACTGAGGTCGGGAGTTCAGACACGACCTGA	165224
Qy	1950	CCACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCTAAGTGTGGTGGCG	2009
Db	165225	CCACATGGAGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCAGCATGGTAGCA	165284
Qy	2010	CATGCGCTGAATCCACGCTACTTTGGGAGGGTGAGGCAGGAGAAATTCCTTGAACCCGGGAG	2069
Db	165285	CATGCGCTGAATCTCAGCTACTCAGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCGGGAG	165344
Qy	2070	GCAGAGGTTGAGTGAACCGAGATCACACCACTGCATCTCCAGCCTGGGGGA - GAGACGGCA	2128
Db	165345	GCAGAGGTTGCAGTGAGCCAAAGATCGCGCCCATTTGCATCTCCAGCCTGGGCAACAAGAGTGA	165404
Qy	2229	GACTTCCTCTCAAAAAACAAAAACAAAAAATTA	2165
Db	165405	AATTTGTACTCAAAAAAACACACAAAAAACAGAA	165441

RESULT 11				
AC005037/c				
LOCUS	AC005037	19508 bp	DNA	linear
DEFINITION	Homo sapiens BAC clone RP11-469M7 from 2, complete sequence.			
				PRI 30-SEP-2000

ACCESSION AC005037
VERSION AC005037.2 GI:4827310
KEYWORDS HTG.

SOURCE	Homo sapiens.
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 190508)
REFERENCE

AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

Access: 6 (12) 100 (1000)
99063792
MEDLINE
PUBMED
9847074
REFERENCE
2 (bases 1 to 100508)

2	(BASES 1. CO 190006)
REFERENCE	
AUTHORS	Abbott, A. and Le, T.
TITLE	The sequence of Homo sapiens BAC clone RPl1-469M7
JOURNAL	Unpublished

BOOKS
REFERENCE
3 (bases 1 to 190508)
Waterston, R. H.
AUTHORS
MIMET
Pittsburgh, Pa.
unpublished

TITLE
 JOURNAL
 Submitted (12-JUN-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 Missouri 63110-1093
 Direct Submission

MO 63108, USA
4 (bases 1 to 190508)
Waterston, R.H.
AUTHORS
REFERENCE

TITLE
JOURNAL
Direct Submission
Submitted (14-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA
REFERENCE
5 (bases 1 to 190508)
AUTHORS
Waterston, R.

TITLE
JOURNAL
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT
On May 14, 1999 this sequence version replaced gi:3309089.
----- Genome Center
Center: Washington University Genome Sequencing Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: gsaia@genome.wustl.edu

Contact: sapriestw@soni.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0469M07

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACes3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-13J8; the clone sequenced to the right is RP11-91M5. Actual start of this clone is at base position 1 of RP11-469M7; actual end is at 190508 of RP11-469M7.

FEATURES

source	Location/Qualifiers
	1..190508
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-469M7"
	/clone_lib="RPCI-11"
repeat_region	155..222
repeat_region	/rpt_family="L2"
repeat_region	982..1299
repeat_region	/rpt_family="Alu"
repeat_region	1502..1533
repeat_region	/rpt_family="(CA)n"
repeat_region	1535..1823
repeat_region	/rpt_family="Alu"
repeat_region	2005..2308
repeat_region	/rpt_family="Alu"
repeat_region	2392..2562
repeat_region	/rpt_family="L1"
repeat_region	3284..3315
repeat_region	/rpt_family="Alu"
repeat_region	3316..3338
repeat_region	/rpt_family="(CAAA)n"
repeat_region	3339..3621
repeat_region	/rpt_family="Alu"
repeat_region	3650..3743
repeat_region	/rpt_family="AT_rich"
repeat_region	4353..4664
repeat_region	/rpt_family="Alu"
repeat_region	4737..5018
repeat_region	/rpt_family="Alu"
repeat_region	5545..5585
repeat_region	/rpt_family="(CATA)n"
repeat_region	5586..5873
repeat_region	/rpt_family="Alu"
repeat_region	5876..6007
repeat_region	/rpt_family="Alu"
repeat_region	6327..6713
	/rpt_family="MER21_g"
	6924..7294
	/rpt_family="MER2_type"
	8027..8049
	/rpt_family="AT_rich"
	8050..8354
	/rpt_family="Alu"
	8355..8661
	/rpt_family="Alu"
	8356..8684
	/rpt_family="AT_rich"
	9450..9752
	/rpt_family="Alu"
	10371..10680
	/rpt_family="Alu"
	10762..11038
	/rpt_family="Alu"
	11039..11109
	/rpt_family="Alu"
	11138..11410
	/rpt_family="Alu"
	11446..11663
	/rpt_family="L1"
	11664..11959
	/rpt_family="Alu"
	11960..12189
	/rpt_family="L1"
	12190..12334
	/rpt_family="Alu"
	12335..12360
	/rpt_family="L1"
	12464..12514
	/rpt_family="L1"
	12515..12822
	/rpt_family="Alu"
	12823..12862
	/rpt_family="L1"
	12857..13096
	/rpt_family="L1"
	13101..13330
	/rpt_family="MaLR"
	13331..13634
	/rpt_family="Alu"
	13635..13692
	/rpt_family="MaLR"
	13704..13990
	/rpt_family="L1"
	14082..14387
	/rpt_family="Alu"
	14402..14704
	/rpt_family="Alu"
	14797..15095
	/rpt_family="Alu"
	15162..15196
	/rpt_family="Mariner"
	15606..15681
	/rpt_family="L1"
	15691..15827
	/rpt_family="MER1_type"
	15861..16189
	/rpt_family="L2"
	16426..16446
	/rpt_family="(CAAT)n"
	16447..16730
	/rpt_family="Alu"
	17072..17179
	/rpt_family="L2"
	17212..17504
	/rpt_family="Alu"
	17869..18139
	/rpt_family="Alu"
	18168..18464
	/rpt_family="Alu"

[illegible]

repeat_region	complement(29311. .29613)	QY	949	GTATTTTATTTGTTG-----AGACAAGGTCTC	1002
repeat_region	/rpt_family="Alu"	Db	30841	GGCAITCAACTCTTGGCTCAAGCAATCCTCGCTT	30782
repeat_region	/rpt_family="Alu"	QY	1003	CACTTCCCAAGTGCATCTTATCTTATCTTATCT	1062
repeat_region	/rpt_family="L1MB7"	Db	30781	ACAGGCATGACCAACCGCAGCCCTCTCTTCCAC	30722
repeat_region	/rpt_family="Alu"	QY	1063	TCCACTTTTGGACCATATGAATAATACCTCCAG	1122
repeat_region	/note="(A)22"	Db	30721	TTCAATTTTCGGGTATATGAATAATGCTATGA	30662
repeat_region	/rpt_type=tandem	QY	1123	GCATATGTTTTCATTTCTGTGTGGTTTATATCT	1182
repeat_region	/rpt_unit=A	Db	30661	GGACATGTTTTCATGTCACAGGTATGTCCTCA	30602
repeat_region	/rpt_family="Alu"	QY	1183	ATAT---TTTGACAGCAGATGTCAGGGGAAGA	1239
repeat_region	/rpt_family="Alu"	Db	30601	AATTCGGTTTGACTGTGTGAGGAGCTGCAAGA	30542
repeat_region	/rpt_family="L1MB7"	QY	1240	GAATCAGCAAGAGTGCAGGG-----TTTTC	1287
repeat_region	complement(32676. .32796)	Db	30541	TACATCTCACAGAGTGTGGAGGGTCTGATTTT	30482
repeat_region	/rpt_family="Alu"	QY	1288	GT-TGACAAATGTGCAGTTTGTATGAAGATA	1346
repeat_region	/rpt_family="Alu"	Db	30481	GTATCACCTGTTTGTGTGTGTTTCTTAAATTT	30422
repeat_region	complement(32836. .32974)	QY	1347	AGGCTGGAATAGGGCTTCAGAGTAAATCATG	1406
repeat_region	/rpt_family="Alu"	Db	30421	AGGAAGTGTACTCACTG-TAGTTTGTGTTTGT	30363
repeat_region	/note="(R)23"	QY	1407	GCTTGCCTCTAAACAAATAATAAATAACACA	1466
repeat_region	/rpt_type=tandem	Db	30362	GTTTTAAATAAATAAATAAATAAATAAATAA	30303
repeat_region	/rpt_unit=T	QY	1467	CTTGCTCTTTTCAACCTGGCTGGAGGAGTGT	1526
repeat_region	complement(34450. .34753)	Db	30302	CTCGCTCTTCGCCAGGCTGGAGTGCAGTGGG	30243
repeat_region	10.0%; Score 415.2; DB 9; Length 4496;	QY	1527	CTCCGGGTTCAAGCAATCTCTCTCTCTCTCT	1586
repeat_region	Best Local Similarity 58.8%; Pred. No. 8.4e-91;	Db	30242	CTCCGGGTTCAACCAATCTCTCTCTCTCTCT	30183
repeat_region	Matches 1044; Conservative 0; Mismatches 633; Indels 99; Gaps 15;	QY	1587	TCCACCATGCCAGCTGAT-TTTTGTATTTTGT	1645
repeat_region	Query Match	Db	30182	CACCAACACCCCGCTAATGTTTGTATTTTGT	30123
repeat_region	428 TAGCTCACTTAACCTCAACCTCCGGCTCA	QY	1646	CCAAGTGTCTCAAACTTTTGTCTCATATTTG	1705
repeat_region	31359 TTGGCAGGCTAGTCTCAACCTCTGAC	Db	30122	CCAGGATGTCTCAAACTCT-----GAACT	30083
repeat_region	488 TAGCTAGGACTACAGCACACAGGCTAC	QY	1706	GGTAGGCCCCCAGACCAAAATAAATCTTAGA	1765
repeat_region	31299 GTGCTGAGATTGATAGCATGACCACT	Db	30082	CTCAGCTCTCCCAAGTGTAGGATTAACAAG	30023
repeat_region	548 TTATGTATTCATTTCTTTTGTGTTGTT	QY	1766	CCACTGTCACTTTGAGAACACACAGTGTGA	1825
repeat_region	31251 TTTTACTGTACTTTGTTTGGGTGGGG	Db	30022	AATGTTTCTTAACAAATAATATAACCCAA	29963
repeat_region	608 TACCCAGGCTGGAGGSCAGTGGCATGTT	QY	1826	CGAAAGAGAAATAGATGAAATAATTTCTC	1885
repeat_region	31191 ACTGGCTGTTTACAGSCACTGTCATGT	Db	29962	TTAAGTCAAAAGACAAACA-----CCCG	29916
repeat_region	668 AGTAGTCTCTCCCTCAGCCTCCCAAGTA	QY	1886	CCAGACATTTTGGAGGCCAAGCATGTGGAT	1945
repeat_region	31135 AGCATCTTCCCACTCAGCTCCCAAGCA	Db	29915	CCAGACATTTTGGAGGCTGAGTGGGTGGT	29856
repeat_region	728 TGGCTGATA-----TTTTTTTCTTCA	QY	1946	CTGGCCCAACATGTTGTAACCCCGTCTCT	2005
repeat_region	31075 TGGCTATTGTCACATCTTTTCTTTTCT	Db	29855	CTGGCCCAACATGTTGTAACCCCGTCTCT	29797
repeat_region	770 CCAGGCTGGATACAGTGGCTGATTAATA	QY			
repeat_region	31016 ACCAGGCTGAAGTGGAGTGGACATCAG	Db			
repeat_region	829 AGCAATCCGCTCGCCTCAGCATCTGAGTA	QY			
repeat_region	30956 AGCAATCTCCCACTCAGCTCCCAAGT	Db			
repeat_region	889 CAGCTAAGTTTAAATAATGATTTTGGTA	QY			
repeat_region	30896 CAGCAATTTTAAAAA-----TTTTTT	Db			

QY 2066 GCGCATCCCTGTATCCAGCTACTTGGGAGGGTGAGCGAGGAGAAATTTCTTGAACCCG 2065
 Db 29796 AGCGGGCATCTGTATCCAGCTACTTCCAGAGGCTGAGCGAGGAGACTTGTGCACTCG 29737
 QY 2066 GGAGCGAGGTTGAGGAGGAGATCACACACTGCATCCAGCTCCAGCTGGGGAGAGAG 2125
 Db 29736 GGAGCGAGGCTGAGTGAGCAAGATCGTGCCACTGCATCCAGCTCCAGCTGGGCAACAGAC 29677
 QY 2126 CGAGCTTCTCTCAAAAAACAAAAAAGAA 2161
 Db 29676 AGAGCGAGCTTCACTCAAAAAAAGAAAAA 29641
 AC006111 190821 bp DNA linear PRI 02-DEC-2000
 DEFINITION Homo sapiens chromosome 16 clone RP11-461A8, complete sequence.
 AC006111
 VERSION AC006111.3 GI:11527451
 KEYWORDS HYG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 190821)
 AUTHORS Ricke,D.O.
 TITLE Large Scale Sequence Analysis and Annotation with the Sequence
 Comparison Analysis (SCAN) System
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 190821)
 AUTHORS Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
 Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
 Goodwin,L., Bryant,J., Tesmer,J., Meinke,L., Longmire,J.,
 White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
 and Deaven,L.
 TITLE Sequencing of Human Chromosome 16p13.3
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 190821)
 AUTHORS Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
 Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
 Goodwin,L., Bryant,J., Tesmer,J., Meinke,L., Longmire,J.,
 White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
 and Deaven,L.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1998) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 REFERENCE 4 (bases 1 to 190821)
 AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
 Bryant,J., Tesmer,J., Meinke,L., Longmire,J., White,S.P.S.,
 Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M.,
 Sutherland,R. and Deaven,L.
 TITLE Direct Submission
 JOURNAL Submitted (02-FEB-1999) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 REFERENCE 5 (bases 1 to 190821)
 AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
 Bryant,J., Tesmer,J., Meinke,L., Longmire,J., White,S.P.S.,
 Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M.,
 Sutherland,R. and Deaven,L.
 TITLE Direct Submission
 JOURNAL Submitted (03-JUL-1999) Center for Human Genome Studies, DOE Joint
 Genome Institute, Los Alamos National Laboratory, MS M888, Los
 Alamos, NM 87545, USA
 REFERENCE 6 (bases 1 to 190821)
 AUTHORS Mundt,M.O.
 TITLE Direct Submission
 JOURNAL Submitted (02-DEC-2000)
 COMMENT On Dec 2, 2000 this sequence version replaced gi:5348385.
 FEATURES Location/Qualifiers

source

1. 190821
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-461A8"
 BASE COUNT 48342 a 48030 c 48379 g 46070 t
 ORIGIN

Query Match 10.0%; Score 415.2; DB 9; Length 190821;
 Best Local Similarity 58.8%; Pred. No. 1.1e-90;
 Matches 1044; Conservative 0; Mismatches 633; Indels 99; Gaps 15;
 QY 428 TAGCTACTGTAACTCAAACTCTCTGGGCTCAAGTGATCTACTACCTCAGCTCCAGCTCCAGAG 487
 Db 163899 TTGGCCAGGCTAGTCTCAAACTCTCTGGGCTCAAGTGATCTACTACCTCAGCTCCAGCTCCAGAA 163958
 QY 488 TAGCTAGACTACAGGACACACAGCCATACCTGCCTAATTTTTTTTTTAATTTTTCAT 547
 Db 163959 GTGCTGAGATTGTAGACATGAGCCACTGCACCGAGCCAGCTTC-----CAAC 164006
 QY 548 TTATGATATTCATTTCTTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 607
 Db 164007 TTTTACTGTAACTTTTGT 164066
 QY 608 TACCAGGCTGGAGGCGAGTGCATGTGACAGCTGACAGGCTTGACTTCTTGGGCTCA 567
 Db 164067 AGTGCTCTTTTACAGGCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 164122
 QY 668 AGTGATCTCTCTGCTCAGCTCCCAAGTAGCTGGGACTACAAACACAGCTGTCAACCATGCC 727
 Db 164123 AGGATCTCTCTCAGCTCCCAAGTAGCTGGGACTACAAACACAGCTGTCAACCATGCC 164182
 QY 728 TGGCTGAGA-----TTTTTTCCTTTTGAACAGGGTATCACTCTGTGTG 769
 Db 164183 TGGCTATTTGCACCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 164241
 QY 770 CCCAGGCTGGAGTACAGTGGGCTAATATAGCTCACTCAGGCTC-CCCTCTCTGGGCTCA 828
 Db 164242 ACCAGGCTGGAATGAGTGGGCTAATATAGCTCACTCAGGCTC-CCCTCTCTGGGCTCG 164301
 QY 829 AGCAATCCGCTGGCTCAGCTCAGCTAGTGTGGGACTACAGGCTGTGCACACAGGCC 888
 Db 164302 AGCAATCTCCCACTCAGCTCAGCTAGTGTGGGACTACAGGCTGTGCACACAGGCC 164361
 QY 889 CAGCTAATTTTAAAAAAGATTTTGTGTATAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 948
 Db 164362 CAGCAATTTTAAAAA-----TTTTTTCGAGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 164416
 QY 949 GTATTTTATTTGTG-----AGACAAGTCTCACTATGTGTGCTATGATCCCTCCACCTC 1002
 Db 164417 GGCAATCAACTCTTGGGCTCAAGCAATCTCTGCTTGGGCTCTCTGAGTGTGTGTGTGTGTGTGTGTGTGTGT 164476
 QY 1003 CACTTCCCAAGTGTCTATCTTATCTGTTCATTTAGTTCAGTTGACAGACATTTAGTGTGTGT 1062
 Db 164477 ACAGGATGACCAACCGCAGCCAGCTCTGTTCACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 164536
 QY 1063 TCCACTTTTGCACATTTAATATATCTCCAGGAAATTTCAATGTATACATTTGTGTGTGT 1122
 Db 164537 TTTCAATTTTCGGGTTTATTAATGAAATATGCTGTATGAACATTTGATGTGTGTGTGTGTGTGTGTGTGTGT 164596
 QY 1123 GCATATGTTTTCATTTCTGTTGGGTTTATATCTAGGAGTGGAAATTTGCTGGATCCCGGTA 1182
 Db 164597 GGCAATGTTTTCATTTCTACAGGTATGTGCTCAGGTTAGAACTGTGGGTCATATGTT 164656
 QY 1183 ATAT---TTTGACAGGAGAGTTTCAGGGGAAAGAAAACTTGGGAAATTAAGACATGTTTA 1239
 Db 164657 AATTCGGTTTTCAGTGTGTGAGGAGCTGCAAGATGTTTTTCTGCGAGAGTTCGACCGTTT 164716
 QY 1240 GAAATCAGCAAGAGTGCAGGG-----TTTTTCGGAGTTTATTTTATTTATTTCT 1287
 Db 164717 TACATTTCTCAGCAGCTGTGGGAGGTTCTGTGATTTTCTCCACATCTAGTTATCCATTTT 164776
 QY 1288 GT-TCACAAATGTGCAGTTTTCATGAGATACAAAGTATATAGTGAGAGAGTGAATTA 1346

Db 164777 GTATCACTGTTTCTTTTGGTGTGTTCTTTAAATTTTAAATATATAGCCACCTTACGAGT 164836
QY 1347 AGGCTGGAATAGGCGGTTCAGAGTAAATCATGAAGCAGCTTTGAATACCAAAATTAAGGA 1406
Db 164837 AGGAAGTGCTAACTCACTG-TAGTTTTCCTTGGTTCCTTAACTCCAGGCACTCTGA 164895
QY 1407 GCTTGGCTGTAAACAAATTAATAAAATCAATTTTTTTTTTTTTTTTTTTTGGAGAAAAGT 1466
Db 164896 GTTTTAAAAAATGATTTAAGAGGAATCTTTTTTTTTTTTTTTTTTTTTTTTGGCGGAGT 164955
QY 1467 CTTCGCTTTTCAACCTGCTGGAGGAGGTGTGATCTCAGCTCACTGCAACTTCGC 1526
Db 164956 CTCGCTGTGCGCCAGCTGGAGTGAGTGCGGGGATCTCGGCTCACTGCAAGCTCCGC 165015
QY 1527 CTCGCGGTTTCAAGCAATTTCTCTGCTTCAGCCCTCCCAAGTAGTGGGACTTACAGGCACT 1586
Db 165016 CTCGCGGTTTCAACCACTTCTCTGCTTCAGCCCTCCCAAGTAGTGGGACTTACAGGCGCC 165075
QY 1587 TCCACCAATGCGGAGCTGAT-TTTTGTATTTTTAGTAGAGTGGGATTTCTACTTTGTGG 1645
Db 165076 CACCACACACCGCGCTAAATGTTTGTATTTTTAGTAGAGCGGGTTTCCACCATGTGG 165135
QY 1646 CCRAGCTGGTCTCAACCTTTTGTGTCATAAATTTGTTCACTATGTTCTCTTTGCTGA 1705
Db 165136 CCAGATGGTCTCAACTCTCT-----GAACTCGTATCCACTTGC 165175
QY 1706 GGTAGGCGCCCGACCAAAAAAATAATCTTAGAATCCAAATCAGTGTGTTGTTTGA 1765
Db 165176 CTCAGCTCTCCCAAGTGTAGATTACAAGCTGAGCCACCGCGCGCTTAAGAGGA 165235
QY 1766 CCATGTCCTAGTGAACACAGTGTGACAGGCTCAGAGTAGAGGTGATCTCTGCT 1825
Db 165236 AATGTTCTTAAACAAATATATAACCCAAAGGCCATAAATATAAGGACCAATATAAAAT 165295
QY 1826 CGAAGAGAAATAGATGAATAATTTCTCCGGCGCAGCGGTGGCTCAGCTGCTGAAT 1885
Db 165296 TTAGTCAAAAGACAACA-----CCGGGTGAGGTGGCTCAGCTGCTGAAT 165342
QY 1886 CCCAGCACTTTGGAGGCGCAAGGATGTGATCACTCAGGTCAGGAGTTCAAAACCCAGC 1945
Db 165343 CCCAGCACTTTGGAGGCGTGAAGTGGTGGTGCCTCAGGTCAGGAGTTCGAGACCCAGC 165402
QY 1946 CTGCGCAATGTTGAACCCGCTCTCTACTTAAATAACAAATTAAGTAAAGTGTGCT 2005
Db 165403 CTGCGCAATGTTGAACCCGCTCTCTACTTAAATAACAAATTAAGTAAAGTGTGCTGCT 165461
QY 2006 GCGCATGCTCTTATCCAGCTACTTGGAGGTTGAGCGAGGAAATTTCTTGAACCCG 2065
Db 165462 AGCGGCACTCTTATCCAGCTACTTCCAGAGGCTGAGCGAGGAGTCTTGCACCTCG 165521
QY 2066 GGAGCGAGGTTTCAGTGAAGCGAGATCACCACCTCAGCTCAGCTGGGGGAGAGAG 2125
Db 165522 GGAAGCGAGGCTGAGTGAAGCGAGATCAGCTCAGCTCAGCTCAGCTGGGCGACAGAG 165581
QY 2126 GAGACTCTCTTCAAAAAACAAAAACAAAAA 2161
Db 165582 AGAGGAGACTCTTACTCAAAAAAATAAAAAA 165617

RESULT 14

AC091180/c AC091180 183690 bp DNA linear PRI 01-MAR-2002
LOCUS Homo sapiens chromosome 17, clone RP11-1079K10, complete sequence.
DEFINITION AC091180
ACCESSION AC091180.20 GI:19033596
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183690)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 17, clone RP11-1079K10
Unpublished
2 (bases 1 to 183690)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Bouckhgalter, B., Brown, A.,
Camarata, J., Campopiano, A., Chang, J., Choepel, Y., Colangelo, M.,
Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K.,
Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D.,
Gindes, B., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,
Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K.,
Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., McPheters, R., Meldrim, J., Meneus, L.,
Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Direct Submission
Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 183690)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Gage, D., Gindes, B., Gindes, S.,
Gindes, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mhova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Direct Submission
Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

4 (bases 1 to 183690)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Gage, D., Gindes, B., Gindes, S.,
Gindes, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mhova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

BglII EcoRI HindIII

	SeqDerMap	FngRPrint	SeqDerMap	FngRPrint	SeqDerMap	FngRPrint
	-----	-----	-----	-----	-----	-----
6121	6298		8696	8797	3998	3805
	-----		-----	-----	-----	-----
2067	2067		6	<800	6382	6481
	-----		-----	-----	-----	-----

```

FEATURES
source      Location/Qualifiers
-----
3775-----3781      642      <800
986-----1023
511-----<800
3242-----3212
2038-----2046
-----
1. .170528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-702D16"
/sex="male"
/tissue_type="blood"
/clone_lib="RPCI human BAC library 11"
complement(17941..18218)
/standard_name="SWSS3783"
43178..43179
159989..160063
/standard_name="SWSS2778"
BASE COUNT 44972 a 35957 c 37542 g 52057 t
ORIGIN

Query Match      9.8%; Score 407.4; DB 9; Length 170528;
Best Local Similarity 58.6%; Pred. No. 8.8e-89;
Matches 1078; Conservative 0; Mismatches 631; Indels 132; Gaps 16;

QY 358 TATCCATATATATATAGACAGGTTCTCAGCTCTGTCACCCAGGCTGGA-GTGAG 416
Db 137212 TATTATTATTATTATTGAGACAGGCTTGTCTGTGTCACCCAGGCTGGAGGTGCAG 137271
QY 417 TGGCACAATATAGTCACTGTAACTCAAACTCCTGGGCTCAAGTGATCCTACTACCTC 476
Db 137272 TAGCATGATCATAGTCACTGAGCTTCAACCTCCTGGCTCAATGATCTCCACCTC 137331
QY 477 AGCCTCAGAGTAGTAGGCTACAGGACACACAGCCATACCTGGCTAAATTTTTTTTT 536
Db 137332 AGCTTCCTGAGTAGTGCGACTGCAAGGATGCACACACGCTGGCTAAATTTTTTTT 137390
QY 537 TAAATTTTCAATTTATGTAATTCATTTCTCTTTTCTGTTGTTGTTTTCAGATAGG 596
Db 137391 -----GTTGTTTGTATTATTGTTAGATAGAG 137419
QY 597 TCTCACTTTGTTACCCAGGCTGGAGGAGTGCGCATGGTGACGTGACGACCTTGACT 656
Db 137420 TTTTACCGTGTGCCAGGCTG-----GTCCTGAAC 137450
QY 657 TCCTGGGCTCAAGTGATCCTCTGCTCAGCCTCCCAAGTAGCTGGGACTACAACACGT 716
Db 137451 TCCTGGGCTCAAGCCTATGCGCAGCTTGGTCTCCAAAGTCTGGGATTACAGGTATAA 137510
QY 717 GTCACCATGCTGGCTGAT-----ATTTTTTTCTTGAACAGGGTATCACTCT 765
Db 137511 GCCACCAACAGGCGCTATTTAATTTTATTATTGCTAGGAGGCTTGC1GT 137570
QY 766 GTTGCCAGGCTGGAGTACAGTGGGTATATAATAGCTCACTGACGCCCTC-CCTCCTGGG 824
Db 137571 GTCACCCAGGCTGGAGTGCAGTGGGCACATATACAGCTCACTGACGCCCTCCTGGG 137630
QY 825 CTCACGAATCCGCTGGGCTCAGCATCCTGAGTACAGTGGGATACAGGCTGTGCGACCA 884
Db 137631 CTCAAAGTGATCTCCTGCTCAGCCTGCTGCTAATAGCTAGGACACAGGTTGTGCGCAA 137690
QY 885 GCCCAGCTAAGTTTAAAAATGATTTTGTATATAGGAGGCTCTGCTATGTTGCTCA 944
Db 137691 TCCCTAGCTAATTTTAAAT-----TTTTTTGTAGAGTGGGATCTCACTATGTTGCTTA 137746
QY 945 GGCTGTATTTTATTGTTGAGACAAAGGCTCACTATGTTGCCATGATCCCCCACTCCA 1004

Db 137747 GGCTGCTTGAACCTCTGGGT-----TCNAGCAGTCCCTTGCCTTGGCTCTCCGA 137799
QY 1005 CTTOCCAAAAGTCTCATCTTAATCTGTTTCATTAGTCAGTTGACAGACATTTAGTTGTTTC 1064
Db 137800 GTGTTGAGATTGCGGGCGCAAGCCACCTCGTCCAACCTGTCTCTTTTGTGTGATGTTAG 137859
QY 1065 CACTTTTTGACCATATATGAATAATFACCTCCAGTGAATATCATGTATACATTTGTGTGGGC 1124
Db 137860 TAACATTTGATGATTTTCTTACTTAGATCCATTAATCATTAAGAGATATACAAATGATGATA 137919
QY 1125 ATATGTTTTCTATCTGTTGGGTTTATATCTAGGAGTGAATTTGCTGGATCCCGGTAAT 1184
Db 137920 TTCTAGTTCTATTATTTCTTTTCTTATTATTAGCTGGAACTTCTA-----TAAA 137967
QY 1185 ATTTTGACAGCAGAGTTTCAGGGGAAGAAACATTTGGGAAATGAAGCATGTTTGAAT 1244
Db 137968 GAGAAACCTCCCTTTTCTAATCTTCGATTTACCCAGTGGTATAATACATATTTGAAAAGT 138027
QY 1245 CAGCAGAGTGCAGGGGTTTTTCGGAGTTTTTATTTATATCTGTTGACAAATGTCAGT 1304
Db 138028 CAGGATGAATATTTTGTCTTTTCCCTTGTTCACCAATTTTCAGAATCTGTATTGGTTC 138087
QY 1305 TTGATGAAGATACAAAGTTATATACTAAGTGAGAGTGAGAAATTAAGGCTGGAATAGGCGCTT 1364
Db 138088 ACTAGCATTCCTCAA-TTAGGCATTGGTTTGTGGGGGTTTTTTTGGTATCATAACTC 138146
QY 1365 CAGAGTAAATCATGAAGCAGCTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAA 1424
Db 138147 TTGGGTTTAAACATTTTGAATATGTTTCAATTCATTCATGCAATTTTTCCTTTCTAGTGA 138206
QY 1425 TAAATAAAAA-----ATCACAAATTTTTTTTTTTTTTTTGGAGAGAGTCTTGCCTTTC 1477
Db 138207 TTTTACTGATATATTTCTTCTTCTTTTCTTTTGTGACGCGAGTCTTGCCTGTT 138266
QY 1478 ACCTTGGCTGGAGGAGTGTGTGATCTCAGCTCACTGCAACTTTTCGCTCCCGGTTTC 1537
Db 138267 GCCCAGGCTGGAGTCAGTGTGTGCAATCTCAGCTAACTGCAACCTCCGCTCTAGGGTTC 138326
QY 1538 AAGCAATTTCTCTGCTTCAGCCTCCCAAGTAGCTGGGACTACAGGCACTTCCACCATGC 1597
Db 138327 AAGTGATTTCTCTGCTCAGTCTCCTGAGTAGTGGGATTCAGGCCATGCAACCTCCG 138386
QY 1598 CCAGCTGATTTTGTATTTTGTAGTAGATGGGATTTCTACTTTGTTGGCAAGAGTGTCT 1657
Db 138387 CCAACT-AAATTTGTATTTTGTAGTAGAGTGGGTTTCTCCATGTTGGTCAGGCTGCT 138445
QY 1658 CAAACTTTT-----TGCTGTCAATAATTTGTTAACTATTGTTTCTTTTGTGTAGGTAGGC 1713
Db 138446 CGAAGCTTTGACCTCAGGTGATCTCCGCTCCGCTCCCAAGTACTAGGATTACAGG 138505
QY 1714 CCCAGACCAAAAAAATAATCTTTAGAAATCCAAATCAGTGTGTTGTTGACCACTGTC 1773
Db 138506 CACCATGCCAGTGGCATATATATTTCTTAATAAATCAGCTTGATGATGAGGACGTG 138565
QY 1774 ACTTGAGAACCAACAGTGTGACCCAGGCTCAGGAGTACAGGTGATCTCTGCTCGAAAGAG 1833
Db 138566 TCTTAGACATCTTCTGTCCCATGTTGCTAGTGGTGGCTTCAATAAATGTCAAATGAA 138625
QY 1834 AATAAGAAAT-----GAAAAATTTCTCCGGGCCAGGCTGGTGGCTCA 1875
Db 138626 TGAGTTGATATTTCTGAGCAGCATTTTAAAAATAAGAGCTAGTGGCTGGGACGCTGCTCA 138685
QY 1876 TGCTGTATATCCAGCACTTTGGGAGGCCAAGCAATGTTGATCACCCTGAGGTCAGGATTT 1935
Db 138686 TGCTGTATATCCAGCACTTTGGGAGGCCAAGCGCGGCGGATCAC--GAGGTCAGGAGAT 138743
QY 1936 CAAAACAGCCTGGCCAAATGTTGAAACCCGCTCTCTACTAAAAATACAAAAAATTTAG 1995
Db 138744 -GAGACCAATTTGGCCCAACATGTTGAAACCCCTGCTCTACTAAAAATAC-AAAAATTAGC 138801
QY 1996 TAAGTGTGTGGCGCATGCCGTAAATCCAGCTACTTGGAGGGGTGAGCAGGAGATTT 2055

```

Db 138802 CGGAGTGGTGGTGCCTTTAGTCCAGCTACTCGGAGGCTGAGCGAGGAATGG 138861
QY 2056 CTTGAACCCGGGAGGCGAGAGTTGCAGTGAAGCGAGATCACACCCTGCACCTCCAGCCTG 2115
Db 138862 TATGAACCCGGGAGGCGAGCTTGCAGTGAAGTTGGCCACTGCACCTCCAGCCTG 138921
QY 2116 GGGGAGAGCGGAGACTTCCTCTCAAAAAACAAAAACAA 2156
Db 138922 GCGACAGAGTGAGACTCTGTCTCAAAAAACAAAAACAA 138962

Search completed: May 20, 2003, 19:44:35
Job time : 18194 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 12:50:07 ; Search time 820 Seconds
(without alignments)
11397.309 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaattcatccattaaat.....tgagtttccagctggccc 4150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseqn-embl/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-embl/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-embl/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-embl/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-embl/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-embl/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-embl/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-embl/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-embl/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-embl/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-embl/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-embl/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-embl/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-embl/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-embl/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-embl/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-embl/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-embl/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-embl/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-embl/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-embl/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-embl/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4150	100.0	4150	24	ABL57445 Human stearoyl-CoA
2	1932.8	46.6	1970	24	ABK94819 Human stearoyl-CoA
3	612.4	14.8	617	22	AAH26627 Human stearoyl-CoA
C 4	429	10.3	17704	23	ABK42978 Genomic sequence #
C 5	427.4	10.3	17707	23	ABK42977 Genomic sequence #
C 6	386	9.3	8880	22	AAK83253 Human immune/haema
C 7	375.4	9.0	92638	22	ABQ88096 Human immune/haema
C 8	368.8	8.9	5491	22	AAK69044 Human immune/haema
C 9	366.4	8.8	9742	22	AAS42015 Genomic sequence #

10	356.2	8.6	84607	20	AAK90847 Human PACAP genom
11	349.8	8.4	12394	22	AAAD14749 Human glycoen syn
12	349.2	8.4	377	21	AAC04246 Human secreted pro
13	349	8.4	13216	22	AAAL05122 Human reproductive
14	349	8.4	13216	23	ABL98014 Human testicular a
C 15	348.6	8.4	396	23	ABV52505 Human prostate exp
16	347.4	8.4	29163	22	AAAL05121 Human reproductive
17	347.4	8.4	29163	23	ABL98013 Human testicular a
C 18	347	8.4	6437	22	AAS42014 Genomic sequence #
C 19	345	8.3	84495	24	AAS20588 Human methionine a
C 20	343.2	8.3	128978	24	ABK83459 Human cDNA differe
C 21	342.8	8.3	26928	20	AAZ32184 Human prothrombin
C 22	342.8	8.3	26928	24	ABN95780 Gene #2278 used to
C 23	337	8.1	109906	24	ABK94411 DNA encoding endot
24	335.2	8.1	128978	24	ABK83459 Human cDNA differe
25	332	8.0	25715	22	AAS33462 DNA encoding human
C 26	329.6	7.9	9372	22	AAAL36684 Human musculoskele
C 27	329.6	7.9	9372	22	AAAL36879 Human immune/haema
C 28	328	7.9	8308	22	AAK68091 Human immune/haema
C 29	328	7.9	8308	22	AAK79820 Human cardiovascular
C 30	327.6	7.9	17397	22	AAS36445 Human cardiovascular
C 31	327.6	7.9	19334	22	AAS36443 Human cardiovascular
C 32	327.6	7.9	19345	22	AAS36444 Human cardiovascular
C 33	327.6	7.9	25423	22	AAK90279 Human digestive sy
C 34	327.6	7.9	25423	22	AAI57656 Human colorectal c
C 35	327	7.9	5932	22	ABA07891 Human ovarian and
C 36	327	7.9	5932	22	AAAL03708 Human reproductive
C 37	326.6	7.9	18010	22	AAK67807 Human immune/haema
C 38	326.4	7.9	32148	22	AAAL04218 Human reproductive
C 39	326	7.9	25424	22	AAK90280 Human digestive sy
C 40	326	7.9	25424	22	AAI57657 Human colorectal c
C 41	325.4	7.8	32145	22	AAAL04631 Human reproductive
C 42	325.4	7.8	32145	23	ABL97544 Human testicular a
C 43	325.2	7.8	24908	22	AAK81665 Human immune/haema
C 44	325.2	7.8	24999	22	AAK65622 Human immune/haema
C 45	325.2	7.8	48727	22	AAK67375 Human immune/haema

ALIGNMENTS

RESULT 1

ABL57445

ID ABL57445 standard; DNA; 4150 BP.

XX ABL57445;

XX ABL57445;

XX ABL57445;

DT 22-AUG-2002 (first entry)

XX Human stearoyl-CoA desaturase gene 5' region.

DE Stearoyl-CoA desaturase; SCD; enzyme; human; promoter; virucide;

KW dermatological; cytostatic; immunosuppressive; antiallergic;

KW antiarthritic; antiinflammatory; cardiovascular; antinaemic;

KW chromosome 10; gene therapy; gene; ds.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FT misc_feature

FT Location/Qualifiers

FT 1..3958

FT /tag= a

FT /note= "region specifically described in Claim 1"

FT misc_feature

FT 1..3910

FT /tag= b

FT /note= "region specifically described in Claim 2"

FT protein_bind

FT complement (105..108)

FT /tag= c

FT /bound_moiety= "HNF-3/Fkh-1"

FT protein_bind

FT 126..129

FT /tag= d

FT /bound_moiety= "SREBP"

FT protein_bind

FT 172..175

FT /tag= e

```
FT /bound_moiety= "NF-Y"  
FT complement (255..258)  
FT /*tag= f  
FT /bound_moiety= "NF-Y"  
FT complement (263..266)  
FT /*tag= g  
FT /bound_moiety= "IRF-2"  
FT complement (461..464)  
FT /*tag= h  
FT /bound_moiety= "SREBP"  
FT complement (541..544)  
FT /*tag= i  
FT /bound_moiety= "IRF-1"  
FT complement (669..672)  
FT /*tag= j  
FT /bound_moiety= "SREBP"  
FT 991..994  
FT /*tag= k  
FT /bound_moiety= "SREBP"  
FT complement (1037..1040)  
FT /*tag= l  
FT /bound_moiety= "AP-1"  
FT complement (1073..1076)  
FT /*tag= m  
FT /bound_moiety= "RAR-alpha"  
FT 1240..1243  
FT /*tag= n  
FT /bound_moiety= "C/EBP-beta"  
FT 1410..1413  
FT /*tag= o  
FT /bound_moiety= "NF-1"  
FT complement (1423..1426)  
FT /*tag= p  
FT /bound_moiety= "HNF3-beta"  
FT complement (1500..1503)  
FT /*tag= q  
FT /bound_moiety= "SREBP"  
FT complement (1587..1590)  
FT /*tag= r  
FT /bound_moiety= "NF-kB"  
FT 1628..1631  
FT /*tag= s  
FT /bound_moiety= "NF-kB"  
FT complement (1632..1634)  
FT /*tag= t  
FT /bound_moiety= "IRF-1"  
FT 1713..1716  
FT /*tag= u  
FT /bound_moiety= "AP-2"  
FT complement (1729..1732)  
FT /*tag= v  
FT /bound_moiety= "HNF3-beta"  
FT 1833..1836  
FT /*tag= w  
FT /bound_moiety= "IRF-2"  
FT complement (1863..1866)  
FT /*tag= x  
FT /bound_moiety= "Whn"  
FT 1917..1920  
FT /*tag= y  
FT /bound_moiety= "SREBP"  
FT 1925..1928  
FT /*tag= z  
FT /bound_moiety= "T3R, RAR-alpha"  
FT complement (2087..2090)  
FT /*tag= aa  
FT /bound_moiety= "Pax-6"  
FT 2093..2096  
FT /*tag= ab  
FT /bound_moiety= "SREBP"  
FT complement (2116..2119)  
FT /*tag= ac  
FT /bound_moiety= "AP-2"  
FT  
FT protein_bind  
FT complement (2179..2182)  
FT /*tag= ad  
FT /bound_moiety= "CHOP"  
FT 2199..2202  
FT /*tag= ae  
FT /bound_moiety= "RAR-alpha"  
FT complement (2231..2234)  
FT /*tag= af  
FT /bound_moiety= "NF-1"  
FT 2245..2248  
FT /*tag= ag  
FT /bound_moiety= "Pax-6"  
FT 2341..2344  
FT /*tag= ah  
FT /bound_moiety= "RAR-alpha"  
FT 2359..2362  
FT /*tag= ai  
FT /bound_moiety= "NF-1"  
FT complement (2365..2368)  
FT /*tag= aj  
FT /bound_moiety= "ER"  
FT complement (2382..2385)  
FT /*tag= ak  
FT /bound_moiety= "NF-1"  
FT complement (2414..2417)  
FT /*tag= al  
FT /bound_moiety= "Rel"  
FT 2445..2350  
FT /*tag= am  
FT /bound_moiety= "AP-1"  
FT complement (2452..2455)  
FT /*tag= an  
FT /bound_moiety= "NF-1"  
FT complement (2533..2536)  
FT /*tag= ao  
FT /bound_moiety= "GR"  
FT complement (2726..2729)  
FT /*tag= ap  
FT /bound_moiety= "CREB"  
FT complement (2744..2747)  
FT /*tag= aq  
FT /bound_moiety= "ER"  
FT 2862..2865  
FT /*tag= ar  
FT /bound_moiety= "NF-1"  
FT complement (2917..2920)  
FT /*tag= as  
FT /bound_moiety= "PPAR-alpha"  
FT complement (2942..2945)  
FT /*tag= at  
FT /bound_moiety= "Oct-1"  
FT 2958..2961  
FT /*tag= au  
FT /bound_moiety= "CREB/ATF"  
FT complement (2965..2968)  
FT /*tag= av  
FT /bound_moiety= "STAT"  
FT complement (2979..2982)  
FT /*tag= aw  
FT /bound_moiety= "C/EBP-beta"  
FT 3056..3059  
FT /*tag= ax  
FT /bound_moiety= "NF-kB"  
FT complement (3062..3076)  
FT /*tag= ay  
FT /bound_moiety= "NF-kB"  
FT complement (3065..3068)  
FT /*tag= az  
FT /bound_moiety= "SP-1"  
FT 3131..3134  
FT /*tag= ba  
FT /bound_moiety= "Ets-1"  
FT complement (3161..3164)  
FT protein_bind
```


QY 3961 AGCGCGACACAGTAGCTGCAAGCGCGCGCTACGCGGTACCGCGGGCTTCG 4020
 Db 3961 AGCGCGACACAGTAGCTGCAAGCGCGCGCTACGCGGTACCGCGGGCTTCG 4020
 QY 4021 AAACCGAGTCTTCGGCGACCGCGAAGTCTCGTCCGGAGCCCTCAGCCCTCGGAAGTG 4080
 Db 4021 AAACCGAGTCTTCGGCGACCGCGAAGTCTCGTCCGGAGCCCTCAGCCCTCGGAAGTG 4080
 QY 4081 ATCCCGCATCCGAGAGCCAGATGCGGCGCACTTGCTCAGGACGATGTGATTTC 4140
 Db 4081 ATCCCGCATCCGAGAGCCAGATGCGGCGCACTTGCTCAGGACGATGTGATTTC 4140
 QY 4141 AGCCTGGCCC 4150
 Db 4141 AGCCTGGCCC 4150

RESULT 2

ABK94819
 ID ABK94819 standard; DNA; 1970 BP.
 XX AC ABK94819;
 XX AC
 DT 29-AUG-2002 (first entry)
 XX XX
 DE Human stearoyl-CoA desaturase (SCD) control region.
 XX XX
 KW Fatty acid regulated gene; polyunsaturated fatty acid disorder;
 KW PUFA disorder; eczema; cardiovascular disorder; hypertriglyceridemia;
 KW dyslipidaemia; atherosclerosis; coronary artery disease;
 KW cerebrovascular disease; peripheral vascular disease; inflammation;
 KW sinusitis; asthma; pancreatitis; osteoarthritis; rheumatoid arthritis;
 KW acne; body weight disorder; obesity; cachexia; anorexia;
 KW psychiatric disorder; cancer; cystic fibrosis; pre-menstrual syndrome;
 KW diabetes; diabetic complication; genetic polymorphism; ds.
 XX XX
 OS Homo sapiens.
 XX XX
 PN WO200240666-A2.
 XX XX
 PD 23-MAY-2002.
 XX XX
 PF 19-NOV-2001; 2001NO-CA01632.
 XX XX
 PR 17-NOV-2000; 2000US-248589P..
 XX XX
 PA (XENO-) XENON GENETICS INC.
 XX XX
 PI Winther MD, Goldberg YP, Knickle LC, Haardt M, Allen SJ, Ponton A;
 PI De Antueno RJ, Jenkins DK, Nwaka SO;
 XX XX
 DR WPI; 2002-508327/54.

Novel isolated polypeptide segment encoded by fat regulated genes,
 useful for diagnosing the presence of or a predisposition for a
 disorder involving fatty acid regulated genes in a subject -
 Claim 23; Fig 26; 225pp; English.
 The invention describes an isolated polypeptide segment (I) whose genes
 are fat regulated. (I) or the polynucleotide encoding it (II) are useful
 for diagnosing the presence of or a predisposition for a disorder
 involving fatty acid regulated genes in a subject. A composition
 containing (I) or (II) is useful for treating a disorder involving fatty
 acid regulated genes, where the disorder is selected from a
 polyunsaturated fatty acid (PUFA) disorder, eczema, cardiovascular
 disorders (such as hypertriglyceridemia, dyslipidaemia, atherosclerosis,
 coronary artery disease, cerebrovascular disease or peripheral vascular
 disease), inflammation (such as sinusitis, asthma, pancreatitis,
 osteoarthritis, rheumatoid arthritis or acne), body weight disorders
 (such as obesity, cachexia or anorexia), psychiatric disorders, cancer,
 cystic fibrosis, pre-menstrual syndrome, diabetes, and diabetic

CC complications. (I) or (II) is useful as research agent and materials for
 CC discovery of treatments and diagnostics for a disease, particularly human
 CC disease. (II) is useful for constructing nucleotide probes and primers,
 CC for detecting genetic polymorphism, for detecting changes in the level of
 CC expression of (II), and as a diagnostic tool. This sequence represents
 CC a control region from a gene encoding a fatty acid regulated protein.
 XX
 SQ Sequence 1970 BP; 428 A; 590 C; 588 G; 364 T; 0 other;

Query Match 46.6%; Score 1932.8; DB 24; Length 1970;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1967; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 2121 GAGAGCAGACTTCTCTCAAAAAACAAAAAAGAAATTAAGCAATTTG 2180
 Db 1 GAGAGCAGACTTCTCTCAAAAAACAAAAAAGAAATTAAGCAATTTG 60
 QY 2181 CAGAGAGAACCTGAAGGGGTCAGACCACGTACAGATTTCTGTCCCATGCCAGTACT 2240
 Db 61 CAGAGAGAACCTGAAGGGGTCAGACCACGTACAGATTTCTGTCCCATGCCAGTACT 120
 QY 2241 TCTGAGGCATGACTGGATGAGCTGTCCACATCTGAAATCATCCAGTCTTGTTCAGAACTT 2300
 Db 121 TCTGAGGCATGACTGGATGAGCTGTCCACATCTGAAATCATCCAGTCTTGTTCAGAACTT 180
 QY 2301 TCACACCGGACAGGGAGCCAGGACTGGAATGCAGTCTCTTGGTCACTGGCCAGAGAGTTG 2360
 Db 181 TCACACCGGACAGGGAGCCAGGACTGGAATGCAGTCTCTTGGTCACTGGCCAGAGAGTTG 240
 QY 2361 GCCTTGACCTGAGACAGTGGCCACAAAGGAGCTGCTTAGTCTACCTCCAGGAATC 2420
 Db 241 GCC-TGACCTGAGACAGTGGCCACAAAGGAGCTGCTTAGTCTACCTCCAGGAATC 299
 QY 2421 CCAGGTCTTGTCTTCTCTGGGAAGTGAATATTTGGCCAGCAGCTCCGTAATTTCTCTCT 2480
 Db 300 CCAGGTCTTGTCTTCTCTGGGAAGTGAATATTTGGCCAGCAGCTCCGTAATTTCTCTCT 359
 QY 2481 TCCCAGGGGAAGGATCTTAGGGCAGTATTTGGGAAGACATATGGGCATGGGAAGGACACGG 2540
 Db 360 TCCCAGGGGAAGGATCTTAGGGCAGTATTTGGGAAGACATATGGGCATGGGAAGGACACGG 419
 QY 2541 GTGAATGCATAGCTGCTGCTGAGCTCTCATGTAGGCTCCTTACGACACGAGAA 2600
 Db 420 GTGAATGCATAGCTGCTGCTGAGCTCTCATGTAGGCTCCTTACGACACGAGAA 479
 QY 2601 AGATGGGGCACAGGACAGATCAGTAGGCTCAGAGCATCTCAGGGACCCGAGGCAATAT 2660
 Db 480 AGATGGGGCACAGGACAGATCAGTAGGCTCAGAGCATCTCAGGGACCCGAGGCAATAT 539
 QY 2661 GGTCTCAGCAGGATTAAGAGCTTGGGCTCTCATATGGTGTCTTCTGGGCTCACTGCCA 2720
 Db 540 GGTCTCAGCAGGATTAAGAGCTTGGGCTCTCATATGGTGTCTTCTGGGCTCACTGCCA 599
 QY 2721 GCTCCGTCACCTTACTGTTGCTGACCATGGCAAGTATTTCATCTCTCATATCTCT 2780
 Db 600 GCTCCATCATTACTGTTGCTGACCATGGCAAGTATTTCATCTCTCATATCTCT 659
 QY 2781 TTCTCTACTTTTAAATGGAATATATGGGGTACCCACCTCCAGGGTCACAGAGAGGTTA 2840
 Db 660 TTCTCTACTTTTAAATGGAATATATGGGGTACCCACCTCCAGGGTCACAGAGAGGTTA 719
 QY 2841 CAGAAAACGATTTCTGGAATTTGGCTTGCAGTATTAATTAATCAATCCTGCCAGTATTCTT 2900
 Db 720 CAGAAAACGATTTCTGGAATTTGGCTTGCAGTATTAATTAATCAATCCTGCCAGTATTCTT 779
 QY 2901 ATTCCACATCCAGCCCTTTTCGCTGCTGCTGGTGAACACATATGTCAGTGTTCCTGA 2960
 Db 780 ATTCCACATCCAGCCCTTTTCGCTGCTGCTGGTGAACACATATGTCAGTGTTCCTGA 839
 QY 2961 CGSTTTTCCCAAGAGATTTCCAAAATTAACAACCTGCCAGTCTGAAGAATCTCCAAAACA 3020
 Db 840 CGSTTTTCCCAAGAGATTTCCAAAATTAACAACCTGCCAGTCTGAAGAATCTCCAAAACA 899

PT Identifying an agent for treating disorders related to serum levels of
 PT fatty acids and glycerols, by contacting screening assay with delta-9
 PT stearyl-CoA desaturase activity with compound and measuring the
 PT activity

XX Claim 12; Page 83; 130pp; English.

CC The present sequence is that of the promoter region of the
 CC human stearyl-CoA desaturase-1 (hscd1) gene. The promoter
 CC structure is similar to that of the mouse SCD1 promoter and
 CC contains conserved regulatory sequences for the binding of several
 CC transcription factors. Like the mouse SCD1, the human SCD1 gene
 CC is regulated by polyunsaturated fatty acids and cholesterol at the
 CC level of gene transcription. The sterol regulatory element binding
 CC protein (SREBP) plays a role in the transcriptional activation of
 CC the gene. The promoter can be used to identify regulatory
 CC elements that modulate and control SCD1 expression in humans, and
 CC to identify regulatory proteins that are suitable targets for small
 CC molecule intervention to modulate expression of SCD1 in humans.
 CC The invention provides screening assays based on the role of hscd1
 CC in human diseases, disorders or conditions relating to serum levels
 CC of triglyceride, very low density lipoprotein, high density
 CC lipoprotein, low density lipoprotein, total cholesterol, or
 CC production of secretions from mucous membranes, monounsaturated
 CC fatty acids, wax esters, etc. Methods for diagnosing and treating
 CC such disorders using modulators of SCD1 activity are also provided.
 CC These disorders include various skin diseases and diverse maladies,
 CC such as cancer and multiple sclerosis, non-insulin-dependent
 CC diabetes mellitus, hypertension, neurological diseases, eye
 CC diseases, and immune disorders as well as cholesterol disorders,
 CC lipidemias and cardiovascular disease.

SQ Sequence 617 BP; 117 A; 211 C; 205 G; 84 T; 0 other;

Query Match 14.8%; Score 612.4; DB 22; Length 617;
 Best Local Similarity 99.8%; Pred. No. 3.1e-114;
 Matches 613; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3459 CCCAGCCCCCTCCAGAGAGAAGCTCCGACGGGGATGCGGCGCAGGCCCGAGCGGG 3518
 DB 4 CCCCCGCCCTCCAGAGAGAAGCTCCGACGGGGATGCGGCGCAGGCCCGAGCGGG 63
 QY 3519 GGTGGAGAGAGCTGAGAGAGAGAAACAGAGGGGAGGGGAGCGAGAGCTGGCGGCAG 3578
 DB 64 GGTGGAGAGAGCTGAGAGAGAGAAACAGAGGGGAGGGGAGCGAGAGCTGGCGGCAG 123
 QY 3579 AGCGAAGACAGAGATTGCGCGAGCCCAATGCAAGCGAGAGCGAGGTGGCAACAAATTC 3638
 DB 124 AGGGAACAGAGATTGCGCGAGCCCAATGCAAGCGAGAGCGAGGTGGCAACAAATTC 183
 QY 3639 CTTGCGCCAATGACGAGCGCGAGTTTACAGAAGCCCTATTAGCATTTCCCGAGAGCGAGG 3698
 DB 184 CTTGCGCCAATGACGAGCGCGAGTTTACAGAAGCCCTATTAGCATTTCCCGAGAGCGAGG 243
 QY 3699 GCGAGGGGAGAGCGCGGGTGTGGTGTGCTGTCGGAGAGATCCCGGCGCCCTGCT 3758
 DB 244 GCGAGGGGAGAGCGCGGGTGTGGTGTGCTGTCGGAGAGATCCCGGCGCCCTGCT 303
 QY 3759 GCGGTGCGCGGAGCTGCGCCCTCTCTCTCCCTCCCGCCCTTACTCTCACCGCGG 3818
 DB 304 GCGGTGCGCGGAGCTGCGCCCTCTCTCTCCCTCCCGCCCTTACTCTCACCGCGG 363
 QY 3819 ACCGCCGCGGCCAGTCAACTCTTCGCACTTTGCCCCCTTGTTCGAGCGGATAAAGGGGG 3878
 DB 364 ACCGCCGCGGCCAGTCAACTCTTCGCACTTTGCCCCCTTGTTCGAGCGGATAAAGGGGG 423
 QY 3879 CTGAGGAATACCGAGACAGGTCAACCGGTTCGAGCTCTAGCCCTTTAAATTCGCCGCTCG 3938
 DB 424 CTGAGGAATACCGAGACAGGTCAACCGGTTCGAGCTCTAGCCCTTTAAATTCGCCGCTCG 483
 QY 3939 GGGACCTCCACGACCGCGGCTAGCGCGGACACACAGCTAGCGTGAAGCGCGCGGCT 3998
 DB 484 GGGACCTCCACGACCGCGGCTAGCGCGGACACACAGCTAGCGTGAAGCGCGCGGCT 543

QY 3999 CAGCGGCTACCGGGCGGCTTCGAAACCGCAGTCTCCGGCGACCCCGAATCCGCTCCGG 4058
 DB 544 CAGCGGCTACCGGGCGGCTTCGAAACCGCAGTCTCCGGCGACCCCGAATCCGCTCCGG 603
 QY 4059 AGCCTCAGCCCCCT 4072
 DB 604 AGCCTCAGCCCCCT 617
 RESULT 4
 ABK42978/c
 ID ABK42978 standard; DNA; 17704 BP.
 XX
 AC ABK42978;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Genomic sequence #877 encoding novel human connective tissue polypeptide.
 XX
 KW Human; connective tissue related disorder; cancer; gene therapy;
 KW cytosstatic; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200155343-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01322.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 25-JUL-2000; 2000US-0220963.
 PR 25-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.


```
Db 17404 AACCGAGTCTCCGCGGACCCGAACTCGCTCCGAGGCTCAGCCCCCTGGAAGTGA 17345
QY 4082 TCCCGGCATCCGAGACCCAGATGCGCGGCCACTTCTCTCGAGCAGCATGTGAGTTTCCCA 4141
Db 17344 TCCCGGCATCCGAGACCCAGATGCGCGGCCACTTCTCTCGAGCAGCATGTGAGTTTCCCA 17285
QY 4142 GCCTGSCCC 4150
Db 17284 GCCTGSCCC 17276

RESULT 5
ID ABK42977/c
XX ID ABK42977 standard; DNA; 17707 BP.
AC ABK42977;
DT DT
XX 21-MAY-2002 (first entry)
XX Genomic sequence #876 encoding novel human connective tissue polypeptide.
XX Human; connective tissue related disorder; cancer; gene therapy;
XX cytosstatic; gene; ds.
XX OS Homo sapiens.
XX WO20015343-A1.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01322.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
```

PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	06-DEC-2000;	2000US-0256719.	
PR	08-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-565190/63.		
XX			
XX	Nucleic acid encoding novel connective tissue associated polypeptides,		
PT	used in diagnosing, preventing, treating or ameliorating a disorder		
PT	such as cancer or rheumatoid arthritis -		
XX			
PS	Disclosure; SEQ ID No 1864; 673pp; English.		
XX			
CC	The present invention relates to the isolation of novel human connective		
CC	tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide		
CC	(cDNA and genomic) sequences encoding them. The sequences of the		
CC	invention are useful in the diagnosis, treatment, prevention and/or		
CC	prognosis of diseases associated with connective tissue(s), including		
CC	cancer. The polynucleotide sequences of the invention are also useful		
CC	in gene therapy. ABK42102-ABK43116 represent genomic sequences encoding		
CC	the novel human connective tissue related polypeptides.		
CC	Note: the sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
XX	Sequence 17707 BP; 4880 A; 4183 C; 4145 G; 4499 T; 0 other;		
XX			
Query Match			
Best Local Similarity	10.3%; Score 427.4; DB 23; Length 17707;		
Matches 428; Conservative	0; Mismatches 1; Indels 0; Gaps		
QY	3722 GTGGTGTCGGTGTCGGAGCATCCCGGGCGGCTGTGTGCGGTCCGCGAGGCTCGGCT	3781	
Db			
17707	GTGGTGTCGGTGTCGACAGCATCCCGGGCGGCTGTGTGCGGTCCGCGAGGCTCGGCT	17648	
QY	3782 CTGTCTCTCCCGCTCCCGCCCTTACCTCCACGGGGACCGCCGGCCAGTCAACTCT	3841	
Db			
17647	CTGTCTCTCCCGCTCCCGCCCTTACCTCCACGGGGACCGCCGGCCAGTCAACTCT	17588	
QY	3842 CGCACTTTGCCCGCTGTGGCAGCGGATAAAGGGGGCTGAGGAATACCGACACGGTC	3901	
Db			
17587	CGCACTTTGCCCGCTGTGGCAGCGGATAAAGGGGGCTGAGGAATACCGACACGGTC	17528	
QY	3902 ACCCGTTGCCAGCTCTAGCTTTAAATTCGCGGCTCGGGGAACTCCACACCGGGCTA	3961	
Db			
17527	ACCCGTTGCCAGCTCTAGCTTTAAATTCGCGGCTCGGGGAACTCCACACCGGGCTA	17468	
QY	3962 GGGCGGACAAACAGTACGCTGCAAGCGCCGCGGCTACAGCGGTACCGGGGGCTTCGA	4021	
Db			
17467	GGCGCGACAAACAGTACGCTGCAAGCGCCGCGGCTACAGCGGTACCGGGGGCTTCGA	17408	

Db 6257 CTTCCAGTACCACTTCTCTCTCTCCCTTCCCTATACCTAATAAACCTTTCTGT

QY	188	CTCTATAAAATTTGCCAATTTCTGGACATTTTCATATAAAATGGGAAGCAAAACAACATGTGAGAC	247
Db			
Db	6317	CTATATAGATTTGGCTTATCTGGACATTTTCATATAAAATGGGAATCAATATATGTGGCCT	6376
QY	248	TTTGTGACTGGCTGGCTTCTCACTTAGCAATTTCTATTTTTAAGGCTCA-TTATGTTTACAGTAC	306
Db			
Db	6377	TTTCTGACTAGCTTCTTTTCACTTAGCGTGATCATTTCAAGGTTCATTTTATGTTGTAGCAT	6436
QY	307	TTAGCACTACTTCAATCTCTTTTATTTCTCAAAATGCTATTCACCTGTGTGGGTATCCCATTA	366
Db			
Db	6437	GTATCACTGTTTGTGTTTGTGTTTGAGACAGGAGTTTGTCT-----	6482
QY	367	TCATATTTATTAGACAGGTTCTCACTCTGTCCACCGAGCTGGAGTGCAGTGGCACAAATC	426
Db			
Db	6483	-----TCTTGCCCAAGCTGGAGTGCAGTGGCGCAATC	6514
QY	427	ATAGCTCACGTGAACCTCAAACTCTCTGGGCTCAAGTGATCTCTACTACTCAGCTCCAGA	486
Db			
Db	6515	TTGGCTCACGTGAACCTCTCTCTCTCTGGTTCCTCAAGTGATCTCTCTGCTCAGCTCCCGA	6574
QY	487	GTAGCTAGGACTACGGCACACACAGCCATACCTCGCTAAATTTTTTTTTTTTAAATTTTCAT	546
Db			
Db	6575	ATAGCTGGGATTAAGAGTGACGCCACCGCACCAATTTTGTACTTTTCAGTACACACA	6634
QY	547	TTTATGATATC-----AATTTCTTTCTTTTTTTGTGTTGTTGTTTGA	591
Db			
Db	6635	TGGGGTTTTCACCATGTTGGCCAGGCTAGTCITTTTTTTTTTTTTTTTTTTTTTTTGTGAGA	6694
QY	592	TAGGGTCTCACTTTGTTTACCGAGCTGGAGGCACTGGCATGGTGACAGCTGA--GCAGC	649
Db			
Db	6695	GGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTGCAAG	6754
QY	650	CTTGACTTCTCTGGGCTCAAGTGATCTCTCGCTCAGCTCCCAAGTAGCTGGGACTACA	709
Db			
Db	6755	CTCGGCTCCCGGTTACAGCCATCTCTCGCTCAGGCTCCCGAATAGCTGGGACTACA	6814
QY	710	AACACGTGTCAACATGCCCTGGCTGATATTT-----TTTTTCTTTGAACAGGGTATCACT	763
Db			
Db	6815	GGCGCTGCCACGCGCCCGCTAATTTTTTAGTATTTTTAGTAGACACGGGTTTCACC	6874
QY	764	CTGTTGCCAGGCTGGAGTACAGTGGCGGTATAATAGCTCACTCGACGCTCCGCTCCTGTG	823
Db			
Db	6875	GTGTTAGCCAGGATGGTCTCGATCTCTCGACTGCTGATCCACCGCTCGGCTCCCA	6934
QY	824	GCCTCAAGCAATCGGCTGGGCTCAGCATCCTGAGTAGCTGGGACTACAGGCTTGTGGCACC	883
Db			
Db	6935	AGTGCTAGTATTACAGGCTTGAAGCCACCGCGCGCGCGCCAGGCTAGTCTTTGAAGTCC	6994
QY	884	AGGCCCAAGCTAAGTTTTAAAAAATGATTTTGGTATAGAGGAGTCTTGTCTATGTTGGTCT	943
Db			
Db	6995	TGACCTCTGGTGATCTGCCCACTCAGCCTCCCAAGTGTGGGATTTACAGGCATGAGCC	7054
QY	944	AGGCTGATTTTTTATGTTGAGACAAGTCTCACTATGTTGCCATGATCCCCCAGCTCC	1003
Db			
Db	7055	ACCATGCGCGGCTGCAGTGTTTTATTCCTTTTGTGCTGAATAGTATCTCATTTTATGG	7114
QY	1004	ACTTCCCAAGTCTCATCTTATCTGTTTCATTAGTCAGTTGTACAGACATTTAGTTGTTT	1063
Db			
Db	7115	ACATAC-----TCATATTTATGCAATTCATTTGTGAGCTAAATAGACATTTGGGTAACTT	7168
QY	1064	CCACTTTTGAACATTAATAATATCTCCAGTGAATATTCATCT---ATACATTTGTGT	1120
Db			
Db	7169	CTACTTTTGGATTAAGAATAATGCTCTGCGCAATTCGTAATCAAGATTTTAAATGT	7228
QY	1121	GGCATATGTTTTCATTTCTGTGGGTTTATATCTAGAGTGGAAATGCTGGATCCCGGG	1180
Db			
Db	7229	GGATCATGTTTTCATTTTCTTGGGTATGACCTAGGAGCAGAAATGATGGGTCAATGTG	7288
QY	1181	TAATATTTTTCACAGCAGGATTCAGGGGAGAGAAAACCTTGGGAAATGAAGCATGTTTAG	1240
Db			
Db	7289	GTGACTCTGTAGAGAGCTTTCAAGCTTGCAGTACTGTTTCCAGAGTGGCTGACATATTTTCC	7348

QY	1241	AAATCAGCAAGATGCGAGGGTTTTTCGGAGTTTATTTTATATTTCTGTCACAAATGTG	1300
DB	7349	ATTGATGTTGAACATCTTTTCATGGCTTATCGACGATTTGTATATCTTTGGAGAAATGC	7408
QY	1301	CAGTTTTCATGAAGATACAAGTTATACTAGTGAGAAAGTGAGAAATTAAGCGTGGAAATAGGG	1360
DB	7409	CCATTCGGTTTCTTGGCGTGATTTTAAATATATTTATATCTTTTATTAATCTAGTGTGAAGCG	7468
QY	1361	CGTTCAGAGTAAAAATCATGAAGCACCTTTCGAATACCAAAATTAAGGAGCTTGGCTGTAAAC	1420
DB	7469	AATTCCTTTAGATAGTCAGATACAAGTCCCTTATCAGA-----TATAT	7511
QY	1421	AAAAATAATAAAATCACAAATTTTTTTTTTTTTTTTGAAGAAGATCTTGTCTCTTCAAC	1480
DB	7512	TATTTGCAAAATATTTTCCCAATTTGTGGGTGTCTTTCGACTTCTTGTCTTTTTCAA	7571
QY	1481	CTGCTCGAGGGCAGTGGTGTATCTACCTTCACTCACACCTTTCGCCTCCCGGGTTCAAG	1540
DB	7572	TTG-----GAGTAAATATACATAACATAAAATTAACATTTTGGGTTTGGGGCAAT	7625
QY	1541	CAATTCCTCGCTTCAGCCTCCCAAGTAGCTGGGACTACAGGCACTTCCCACCATGCCA	1600
DB	7626	GGTTCACACCTGTAAATCCACGAAATTAGGAGGGCAGGGTAGGAGGATGTCTCAGCCCA	7685
QY	1601	GCATGATTTTGTATTTTATAGTAGAGATGGATTTTCACTTTGTGGCCAAAGCTGTCTCAA	1660
DB	7686	GGAGTTTGGAGACCAGTCTGGGCAACATACGAAGAATCTGTCTACAAA-----	7734
QY	1661	ACITTTTTCGTCTATAATTTGTGTAACATATGTTCCTTTTGTCTGAGGTAGGGCCCCCAGA	1720
DB	7735	-----ATATTTAAAAATTAGCTGGGGCTGGCTGCACACTTCTAGTCCCAAGGA	7781
QY	1721	CCAAAAAATAAATCTTAAATCCAAATCAGTGTGTGGTTTGACCACCTGTCACTTGG	1780
DB	7782	TGAGGTGAGAAATCCCTTGAGTCCAGGAGGTCAAGGCTGTAAGTGAGCCGTGATCACAC	7841
QY	1781	AACCAAGTGTGACCAAGGCGCTCAGGAGTAGAGTGTATCTCTGTCGAAAGAGAAATAGA	1840
DB	7842	CAGTGCAGTCCAGCCTG-----ATGGCAGAGTGAGACTCTGTATTTTTTAAAAA	7894
QY	1841	ATGAAATATTTCCGGGCCAGCGTGGTGGCTCATGCTGTATATCCAGCACTTTGGGA	1900
DB	7895	AACAAAAACAAAAGAGCGCAGCGGGTGGCTCAGCCTGTAACTCAAGCCTTTGGGG	7954
QY	1901	GGCCAAGGATGTGGATCACCTGAGTCAAGGATTCAAAACCGCTGGCCACACATGGTG	1960
DB	7955	GACCAAGGCGGAGATCACTGAGTCCGAGGTCGAGACCAAGCTTACCTACACGGAG	8014
QY	1961	AAACCCCGTCTCTACTAAAAATACAAAATATAGCTAAGTGTGGTGGCGCATGCCGTAA	2020
DB	8015	AAACCGTGTCTCAACTGAAAAATTC--AAAAATTGCCAGGTGTGGTGGCGCATGCTGTAA	8072
QY	2021	TCCCAGCTACTTGGGAGGTGAGGCAGGAGAAATTTCTTGAACCCGGGAGCGCAGGTTGC	2080
DB	8073	TCCCAGCTATTCAGGAGGTTGAGGCAGGAGAAATCGTTTGAACCTTGGGAGCGCAGGTTGC	8132
QY	2081	AGTGAAGCGAGATCACACCACCTGCACCTCCAGCTGGGGGA--GAGAGCGGACACTTCCTCTC	2139
DB	8133	GGTGAAGCCAAAGATCGCACTTGTCACTTAGCCTTGGGCAACAGAGCGCAACTCCGCTCTC	8192
QY	2140	AAAAAACAAAAACAAAGAA	2161
DB	8193	AAAAAAAAAAAAAAAAAGAAAA	8214

RESULT 7	
ABQ88096/c	
ID	AB288096 standard; cDNA; 92638 BP
XX	
AC	ABQ88096;
XX	
DT	18-SEP-2002 (first entry)
XX	

Db	56634	TAGAGACGGGGTTTACCTTGTTAGCCAGGATGGTCTCGGACTCTGACCTTGTGTGATCGG	56575	PR	11-JUL-2000;	2000US-02174819
Qy	1670	----TGTCAATAATTGTTGAACATAATTGTTTCCTTTTCTGCTGAGGTAGGGCCCCCAGACCCAA	1725	PR	11-JUL-2000;	2000US-02174990
Db	56574	CCGGCTCGACCTCTCAAGTCTGGGATACAGCGGTAGCCACCGCCACCGGCGAGTTT	56515	PR	26-JUL-2000;	2000US-02209633
Qy	1726	AAAAATAATCTTAGAATCCAAATCAGTGTGTGTTTGACACACTGTCTCACTTGAGAACC	1785	PR	26-JUL-2000;	2000US-02209644
Db	56514	ATTTGGAATCTCTCGATGGGAGATTTGGCTATTCTCCCTATTATTCATTCATCAT	56455	PR	14-AUG-2000;	2000US-02245119
Qy	1786	CAGTGTGACCGGCTCAGGAGTAGAGTGATCTCTGCTGAAAGAGAAATAGAAATGAA	1845	PR	14-AUG-2000;	2000US-02245133
Db	56454	TTATATCAGCATGGACTTACAGGTATTTATTTATACCTTTGGATATGATCTCAAACTAC	56395	PR	14-AUG-2000;	2000US-02252114
Qy	1846	ATAATTCCTCC-----GGGCCAGGCGTGTGGCTCATGCTCTAATCCAGCACTTT	1896	PR	14-AUG-2000;	2000US-02252166
Db	56394	TTAATTTTGTGCCAAGTGTGCCAGGCGTGTGGCTCACACCTGTAGTCCAGCACTTT	56335	PR	14-AUG-2000;	2000US-02252677
Qy	1897	GGGAGCCAGGCGGATGCACCTCAGGCTCAGGAGTTCAAAACCCAGCTTGCCCAACAT	1956	PR	18-AUG-2000;	2000US-02262759
Db	56334	GGGAGCCAGGCGGATGCAC--GAGTCAAGGATCAGACCATCTCTGGCTTAACAC	56277	PR	22-AUG-2000;	2000US-02262799
Qy	1957	GGTGAACCCCGTCTCTACTAAATAACAAAATAGTCTAGTGTGGCGCATGCT	2016	PR	22-AUG-2000;	2000US-02266868
Db	56276	AGTGAACCTTGTCTCTATTAATAACAAAATATAGCCAGCATGTTGGCGCGCGCT	56217	PR	23-AUG-2000;	2000US-02271822
Qy	2017	GTAATCCAGCTACTTGGGAGGGTGAGCAGAGAGATTTCTTGACCCGGGAGGCGAGG	2076	PR	30-AUG-2000;	2000US-02277009
Db	56216	GTAGTCCAGCTACTTGGAGGCTGAGCAGGAGAGTGGTGTGACCCCGGAGGCGGAGC	56157	PR	01-SEP-2000;	2000US-02289243
Qy	2077	TTCAGTGAAGCGAGATCACACCTGACACTGCCAGCTCCAGCTTGGGGAGAGCGAGACTTCT	2136	PR	01-SEP-2000;	2000US-02292877
Db	56156	TTCAGTGAAGCGAGATCGGCCACTGC-CTCCAGCTTGGGTGACAGAGCGAGACTCCAT	56098	PR	01-SEP-2000;	2000US-02293443
Qy	2137	CTCAAAAAACAAAAACAAAAGATT	2163	PR	01-SEP-2000;	2000US-02293455
Db	56097	CTCAAAAAAGTAAAAATAAAATTT	56071	PR	05-SEP-2000;	2000US-02295059
RESULT 8					05-SEP-2000;	2000US-02295133
AAK69044/c					06-SEP-2000;	2000US-02304377
ID	AAK69044 standard; DNA; 5491 BP.				08-SEP-2000;	2000US-02312443
XX					08-SEP-2000;	2000US-02314113
AC					08-SEP-2000;	2000US-02320801
CT					08-SEP-2000;	2000US-02320811
DT					12-SEP-2000;	2000US-02319686
XX					14-SEP-2000;	2000US-02323397
XX					14-SEP-2000;	2000US-02323398
XX					14-SEP-2000;	2000US-02323399
XX					14-SEP-2000;	2000US-02324001
XX					14-SEP-2000;	2000US-02324011
XX					14-SEP-2000;	2000US-02330633
XX					14-SEP-2000;	2000US-02330644
DE					21-SEP-2000;	2000US-02330655
XX					21-SEP-2000;	2000US-02342233
XX					25-SEP-2000;	2000US-02342744
KW					25-SEP-2000;	2000US-02349977
KW					26-SEP-2000;	2000US-02349988
XX					27-SEP-2000;	2000US-02354484
XX					27-SEP-2000;	2000US-02358334
XX					29-SEP-2000;	2000US-02363277
PN	W0200157182-A2.				29-SEP-2000;	2000US-02363677
PD					29-SEP-2000;	2000US-02363688
XX					29-SEP-2000;	2000US-02363699
PF						

20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 23856; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting

CC the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, CC diagnose and treat immune/hematopoietic-related diseases, especially, CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703 CC to AAK87694 represent human immune/hematopoietic antigen genomic CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169 CC represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 5491 BP; 1662 A; 1226 C; 1439 G; 1164 T; 0 other;

Query Match 8.9%; Score 368.8; DB 22; Length 5491;
Best Local Similarity 56.0%; Pred. No. 7.1e-65;
Matches 1023; Conservative 0; Mismatches 627; Indels 178; Gaps 10;

QY 361 CCCATATCATATATATAGAGACAGGTTCTCAGTCTGTACCCAGGCTGGAGTGCAGTGC 420

Db 3658 CCTAATTAATTTTGTGAGACGGTGTCTCACTCTCTCCCAAGTTGAGTGAATGAC 3599

QY 421 ACAATCATAGCTCACTGTAACTCAAACTCTGGGCTCAAGTGAATCTACTACTCAGCC 480

Db 3598 TGAATCTTGGCTCACTGCAACTCCACCTACCGGTTCAAGTGAATCTCCCGCTCAGCC 3539

QY 481 TCCAGAGTAGTAGTACAGGACACAGCCATACCTGCTGCTAAATTTTATTAAT 540

Db 3538 TCCCAAGTAGCTGGGACTATAGCGCGCACCATCAGCTTGGCTAATTTTGTATT--- 3483

QY 541 TTTCATTTTATGATTCATTTCTTTCTTTTGTGTTGTTTGTGATAGGCTCTC 600

Db 3482 -----TTTAGTAAAGACAAGGTTTT 3463

QY 601 ACTTTGTTTACCAGGCTGGAGGCGAGTGGCATGGTGACAGCTGAGCAGCTTGACTTCCT 660

Db 3462 GCATGTTGGCCAGGCTGCT-----CTTGAATCAT 3432

QY 661 GGGCTCAAGTAGTCTCTGCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGTGTC 720

Db 3431 GACCTCAGGTGATCGGCTCACATCAGCCTCCCAAGTGTGGGATTACAGGCTGAGCCA 3372

QY 721 CCATGCTGGCTGATA-----TTTTTTTCTTGAACAGGGTATCA 761

Db 3371 CTGTGCTCTGCCATTAATTTATTTATTTATTTATTTTATTTTATTTAGCAGAGTCTCA 3312

QY 762 CTCTGTGCCAGGCTGGAGTACAGTGGCGTAATATAGCTCACTGAGCCTCC-CTCTCC 820

Db 3311 CTCTGTGCCAGGCTGGAGTACAGTGGCGTAATATAGCTCACTGAGCCTCCATCTCC 3252

QY 821 TGGGCTCAAGCAATCGGCTGGGCTCAGCATCTCTGAGTACCTGGGACTAGAGCTTGCTCC 880

Db 3251 TAGGTTCAAGCTATTCTCTGCTCAGCCTCCCAAGTAGTGGGATTACAGGCGCCACC 3192

QY 881 ACCAGGCCAGCTAAGTTTAAATAATGATTTTGTATAGAGGAGTCTTGTATGTTG 940

Db 3191 ACCACACCGGCTAATTTTGT-----ATTTTGTAGATAGAGTTCACCAATGTTG 3138

QY 941 CTCAGGCTGATTTTATTTGTTGAGACAGGCTCACTATGTTGCCATGATCCCCCACC 1000

Db 3137 GCCAGSGGGTCAACCACTCTCGGCTCA-----ATCATCCACCTGCC 3094

QY 1001 TCCACTTCCCAAGTGTCTATCTTATCTGTTTCAATGATGAGTTCAGACACATTTAGTTG 1060

Db 3093 TCAGGCTCCCAAGTGTGGGATTACAGGCGTGATCCACCGCGCGCGGCTGAATGTTTC 3034

QY 1061 TTTCCACTTTTTCACCATATGAATACTCCAGTGAATATTCATGATATGATTTGTT 1120

Db 3033 TTTCTTTCTTTTATTTTCTTTTATTTAGAGACATAGCTTGTCTGTTTACCAATGCT 2974

QY 1121 GGGCATATGTTTCTATTTCTGTTGTTTATATCTAGGAGTGAATTTGCTGATCCCGG 1180

Db 2973 GGAGTACGGTGGGAGATCAGCACTCACTACAACTTGAATTCCTGGATTCAAGCAATCC 2914

QY 1181 TAATATTTGACAGGACAGATTCAGGGGAGAGAAAACCTTGGGAAATGAAGCATTTTAT 1240

Db 2913 TCCCATCTCAGCTCTCTGAGT-----AGCTAAG 2886

```
QY 1241 AATACAGAGAGTGTGAGGGGTTTTTCGGAGTATTTATTTATATTCGTGTTGACAAATGTG 1300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2885 GACTACAGGTGCGTATACACACACCTCGCTGCTGTTCTGTAATATCTGTTTATATAACTCC 2826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1301 CAGTTTATGATCAAGATCAAGATTTACTAAGTGAAGTGAAGATTAAGGCTGGAATAGGG 1360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2825 CTATGTTGTTCAATACCTCTCTCAAGAAATACTATTTATAGTCTGCTCTCCGGAATTT 2766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1361 CGTTCAAGATTAATCATGAGACACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAC 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2765 TCTTGATCTTCAACTGAATGTTTCTGTATCTCTGGGACTGAATTTCTT----- 2712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1421 AATAATAAAAAATCACAATTTTTTTTTTTTTTTTGGAGAAAGAGCTTCTCTTTTCAAC 1480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2711 -----TTCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGATACTTCTCTGTCAAC 2665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1481 CTGGCTGAGGGCAGTGTGTGATCTCAGCTCACTGCAACTTTCGCCTCCCGGTTTCAAG 1540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2664 CAGGCTAGAGTTCAAGTGGCTTGATCTCAGCTCATTTGCAACTCCGCTCTCGGTTCAAG 2605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1541 CAATCTCTCTTCAAGCTTCCCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2604 CAATCTCTCTCTCAAGCTTCCCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1601 GCTGATTTTTGTATTTTACTGAGATGGGATTTCACTTTTGTGGCCAAAGCTGTCTCAA 1660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2544 ACTAATTTTTGTATTTTACTGAGATGGGATTTCACTTTTGTGGCCAAAGCTGTCTCAA 2485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1661 ACTTTTGTCTCATATTTGT-----TGTAATATTTGTCTTTTGTCTGAGGTAG 1710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2484 ACTCTGACCTCAAGTGTGCTGCGGCTCGGCTCCCAAGTGTCTGGATTTATAGSCAT 2425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1711 GGCCCCAGACCAAAAAAATAATCTTAGAATCCAAATCAGTGTGTGTTGTTGACCACT 1770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2424 GAGCCACCATGCCAGCCATGAGACTGAATTTCTGCTGTGTTTGTGTTTGTGGGCTCTATT 2365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1771 GTCACTTGTAGAACACAGTGTGACAGGGCTCAGGAGTGTGATCTCTGCTCGAATA 1830
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2364 TTATGTTAGGACTGTTTCTCCACGGTTTGTGTATCTTTGGTCTATATTAATCTCTACA 2305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1831 GAGAAATAGAAATATAATCTCCGGCCAGGGTGTGCTCATGCTGTAATCCAG 1890
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2304 AGCAAGCACTAAAGCTGTTGAAGCCAGACACAGCGCTCAAGCTGTAATCCAG 2245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1891 CACTTTGGAGGCAAGGATGTGGATCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2244 CACTTTGGAGGCTGAGGCGAGGATCA--TGAGGTCAGGAGTGTGAGTGTGAGTGTGAGTGTG 2187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1951 CAACATGTTGAACCCGCTCTCTACTAAAAATCAAAAAATTAGCTAAGTGTGTTGGCGC 2010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2186 TAACCTGGTGAACCCCTCTCTACTAAAAATCAAAAAATTAGCTAAGTGTGTTGGCGG 2127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2011 ATGCTCTTAATCCAGTACTTGGAGGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2070
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2126 GTGCTCTAGTCTCAGTACTTGGAGGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2067
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2071 CAGAGTGTGAGTGAACGAGATCACACCCTGCTCTGAGTGTGAGTGTGAGTGTGAGTGTG 2130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2066 CGGAGCTTGCAGTGTGAGGAGATTTATCCACTGCTCTGAGTGTGAGTGTGAGTGTGAGTGTG 2007
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2131 CTCTCTCTCAAAAAAACAACAAAAA 2158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2006 CTCTCTCTCAAAAAAACAACAAAAA 1979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 9

AAS42015/c

ID AAS42015 standard; DNA; 9742 BP.

XX

AC AAS42015;

XX

DT 17-DEC-2001 (first entry)

XX Genomic sequence #331 encoding novel human enzyme polypeptide.

DE Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritis; nephrotoxic; anticoagulant; ds.

XX Homo sapiens.

XX WO200155301-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01239.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0217496.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226686.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 05-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229513.

XX 06-SEP-2000; 2000US-0230437.

XX 08-SEP-2000; 2000US-0230438.

XX 08-SEP-2000; 2000US-0231242.

XX 08-SEP-2000; 2000US-0231243.

XX 08-SEP-2000; 2000US-0231244.

XX 08-SEP-2000; 2000US-0231413.

XX 08-SEP-2000; 2000US-0231414.

XX 08-SEP-2000; 2000US-0232080.

XX 08-SEP-2000; 2000US-0232081.

XX 12-SEP-2000; 2000US-0231968.

XX 14-SEP-2000; 2000US-0232397.

XX 14-SEP-2000; 2000US-0232398.

XX 14-SEP-2000; 2000US-0232399.

XX 14-SEP-2000; 2000US-0232399.

QY	370	TATTATTAGAGACAGGTTCTCACTC-TGTCACCCAGGCTGGAGTGCAGTGGSCAACATCAT	428
Db	6351	TTTTCTTTGAGATGGAGTTTCACTCTGTGTGCCAGGCTGGAGTGCCTGTGCGCGTACTCTT	6292
QY	429	AGTCTACTGTAACTTCAAACTCCTGGGCTCAGTGATCCTACTACCTCAGCCTCCAGAGT	488
Db	6291	GGCTATTGCAACTCTCGGCTCCGGGGTTCAAGCAATTCCTCTGCCTCAGCCTCCTCAGT	6232
QY	489	AGCTAGGACTTACAGGCACACACAGGCATACCTTGGCTAAATTTTTTTTTTAAATTTTCATTT	548
Db	6231	AGCTGGGGTTACAGGCATGTGCCACACGCCCGGCTAAATTTGTATTTTTTA-----	6181
QY	549	TATGTATTCAATTTCTTCTCTTTTTTTTGTGTGTGTGTGTGTAGATAGGCTCTCACTTTGTT	608
Db	6180	-----GTAGGACGAGGGTTTCTCCANGTT	6157
QY	509	ACCCAGGCTGGAGGCGAGTGGCATGGTGACAGCTGAGCAGCCTTGACTTCCTGGGCTCAA	668
Db	6156	GGTCAGGCTGGT-----CTTGAACCTCCGACCTCAG	6126
QY	569	GTGATCCTCCTGCCTCAGCCTCCCAAGTAGCTGGACTTACAAACAGCTGTCAACATGSCCT	728
Db	6125	GTGATCCGGCCGCTCGGCTTCCCAAAGTGTGGGATTTTACAGGCGTGAGCCACAGCGCTC	6066
QY	729	GGCTGATATTTTTTTCTTGAACACAGGGTATCACATCTGTGCCACAGGCTGGAGTACAGTG	788

XX Use of a gene linked to the region of a chromosome associated with
PT neuropsychiatric disorders, especially for diagnosis of bi-polar
PT affective disorder and schizoaffective disorder manic type -
PS Disclosure; Fig 2A-2U; 99pp; English.
XX
XX The present genomic sequence is that of human pituitary adenylate cyclase
CC activating polypeptide (PACAP). This gene is localised to the short arm
CC of chromosome 19p11 and synthesised in the retinal ganglion cells. The
CC genomic DNA is isolated from lymphocytes of individuals. The coding
CC region of the gene has a dinucleotide repeat polymorphism marker known
CC as W3440 and an STS marker, ADCYAP1 is located in the 3'UTR. These
CC markers are useful to map the gene physically and genetically to the
CC chromosome 18. This sequence can be used to design probes for detecting
CC the presence or absence of a mutation in the wild type PACAP gene. It is
CC also useful for genetic diagnosis of neuropsychiatric disorders like bi-
CC polar affective disorder (BP-1) and schizoaffective disorder manic type
CC (SAD-M), which are more valid than current clinical diagnosis based on
CC clinical observation and subjective reports. It is used for identifying
CC compounds that modulate PACAP gene expression which are useful for
CC treatment of PACAP disorders.
XX
XX Sequence 84607 BP; 23055 A; 17765 C; 18980 G; 24744 T; 63 other;

Query Match 8.6%; Score 356.2; DB 20; Length 84607;
Best Local Similarity 68.6%; Pred. No. 4.4e-62;
Matches 522; Conservative 0; Mismatches 233; Indels 6; Gaps 2;
QY 1429 AAAAAATCACAATTTTTTTTTTTTTTTGAGAAAGAGCTTGCTCTTCACCCCTGCGTGG 1488
DB 20031 ACAACTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 20090
QY 1489 AGGCGAGTGGTGTATCTCAGCTCCTACTGCACTTTTCGCTCCCGGGTTCACGCAATTC 1548
DB 20091 ACTACAATGGCATGTCTCAGCTCCTACTGCACTCTGCTCCCGGGTTCACGCAATTC 20150
QY 1549 CTGCTTCAGCTCCCAAGTACTGAGTGGAGTACAGGCACTTCCACCACGCCAGCTGATT 1608
DB 20151 ATGCTCTCAGCTCCCGAGTGGAGTACAGGCTGCTGGGCTGGCCAGCCACCCAGCTA 20210
QY 1609 TTGATTTTTAGTAGAGTGGATTTCTACTTTGTTGGCAAGCTGGTCTCAAACTTTTG 1668
DB 20211 TTGATTTTTAGTAGAGTGGGTTTACCAGTTGTTGGCAGGCTGTCTCGAAC---TC 20266
QY 1669 CTGTCATATTTGTTGTAATTTGTTTCTTTGCTGAGTAGGGCCCCCAGACCAAAAAA 1728
DB 20267 CTGACCTCGTGGTTCGCCATCTCGGCCCTCTCAAAAGTGTGGGATTACAGGTGTG 20326
QY 1729 AATAAATCTTAGAATCCAAATCAGTGTGTGTTTGGACCACTGTCTACCTTGAGAAC 1788
DB 20327 CGTGCTTGGCTTACAAATTTGATTTAGAGGTGATAGATGATTATTATTGCTGTGT 20386
QY 1789 TGTGACAGGGGCTCAGGAGTAGAGTGTATCTCTCTCGAAAGAGAAATAGAAATGAA 1848
DB 20387 TCTAGTTGTATCCAAAAAATCAAAATTAAGAGTTAAACCAAAATAGCACATTTA 20446
QY 1849 ATTCTCGGGGCGAGGCTGGTGGCTGTCCTGTAATCCAGCACTTTGGGAGGCCAAG 1908
DB 20447 ATCATACGGGCGGGCGGGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGGCCG 20506
QY 1909 CATGTGGATCACCCTGAGTCCAGGAGTTCAAAACACGCTGGCCCAACATGGTGAAC 1968
DB 20507 TGGCGGGATCAC--GAGGTTCAGGAGATCGAGACCATCTCTGGTAAACGGTGAAC 20564
QY 1969 TCTCTACTAAAAATCAAAAAATTAGCTAAAGTGTGGTGGCGCATGCTGTATCTCCAG 2028
DB 20565 TCTCTACTAAAAATCAAAAAATTAGCGGGGTAGTGGCGGGCGCTCTCTCCAGCT 20624
QY 2029 ACTTGGGAGGTTAGCGAGGAGATTTCTTGAACCCGGGAGCGAGGTTGCAAGTGAAG 2088
DB 20625 ACTCGGAGGCTGAGGAGGAGATGGGTGAACCCGGGAGCGGAGCTTGCAGTGAGCC 20684

QY 2089 GAGATCACCACTGCACCTCCAGCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2148
DB 20685 AAGATCGGCCACTGCCTCCAGCTGGCGGACAGAGAGAGAGAGAGAGAGAGAGAG 20744
QY 2149 AAAAAACAAAGAAATTAAGCAAAATTAGACATTTGCAGATTCAGAGAGAA 2189
DB 20745 AAAAAAAGAAAAAAGGAAATGATCATCTTGGAAAGGAA 20785
RESULT 11
AAD14749
ID AAD14749 standard; DNA; 12394 BP.
XX AAD14749;
XX 01-NOV-2001 (first entry)
XX Human glycogen synthase kinase 3 alpha genomic DNA.
XX Human; glycogen synthase kinase 3 alpha; antidiabetic; cytostatic;
KW antitense therapy; diabetes; hyperproliferative disorder; inflammation;
KW neurological disorder; tumour; haematopoietic disorder; infection;
KW hyperproliferative disorder; developmental disorder; ds.
XX HOMO sapiens.
XX FH Key Location/Qualifiers
FT CDS 115..11786
FT FT /product= "Human glycogen synthase kinase 3 alpha"
FT FT /tag= a
FT FT 115..397
FT FT /tag= b
FT FT 398..2437
FT FT /tag= c
FT FT 2438..2625
FT FT /tag= d
FT FT 2626..5638
FT FT /tag= e
FT FT 5639..5722
FT FT /tag= f
FT FT 5723..5863
FT FT /tag= g
FT FT 5864..5974
FT FT /tag= h
FT FT 5975..7901
FT FT /tag= i
FT FT 7902..8032
FT FT /tag= j
FT FT 8033..8120
FT FT /tag= k
FT FT 8121..8227
FT FT /tag= l
FT FT 8228..9196
FT FT /tag= m
FT FT 9197..9294
FT FT /tag= n
FT FT 9295..9374
FT FT /tag= o
FT FT 9375..9470
FT FT /tag= p
FT FT 9471..9897
FT FT /tag= q
FT FT 9898..10084
FT FT /tag= r
FT FT 10085..10430
FT FT /tag= s
FT FT 10431..10523
FT FT /tag= t
FT FT 10524..11712
FT FT /tag= u
FT FT 11713..11786
FT FT /tag= v
XX

PN WO200152865-A1.
 PD 26-JUL-2001.
 XX 16-JAN-2001; 2001WO-US01411.
 XX 21-JAN-2000; 2000US-0488856.
 XX (ISIS-) ISIS PHARM INC.
 XX Monia BP, McKay R, Butler MM, Wyatt JR;
 XX WPI; 2001-442247/47.
 DR P-PSDB; AAE06521.
 XX
 PT Antisense compound 8 to 30 nucleobases in length comprising a compound
 PT that is targeted to a nucleic acid molecule encoding glycogen synthase
 PT kinase 3 alpha, useful for the treatment of e.g. diabetes and
 PT hyperproliferative disorders -
 XX
 PS Example 15; Page 92-100; 115pp; English.
 XX
 CC The invention relates to an antisense compound 8 to 30 nucleobases in
 CC length targeted to a nucleic acid encoding glycogen synthase kinase 3
 CC alpha. The antisense compound specifically hybridises with and inhibits
 CC the expression of glycogen synthase kinase 3 alpha. The antisense
 CC compound is useful for the treatment of a diseases associated with
 CC glycogen synthase kinase 3 alpha such as diabetes, a neurological
 CC disorder, a haematopoietic disorder, a hyperproliferative disorder or
 CC a developmental disorder. The antisense compounds may also be used
 CC prophylactically to prevent or delay infection, inflammation or tumour
 CC formation. The present genomic DNA sequence encodes human glycogen
 CC synthase kinase 3 alpha protein.
 XX
 SQ Sequence 12394 BP; 2780 A; 3043 C; 3547 G; 3024 T; 0 other;
 Query Match 8.4%; Score 349.8; DB 22; Length 12394;
 Best Local Similarity 55.2%; Pred. No. 5.7e-61;
 Matches 1016; Conservative 0; Mismatches 742; Indels 81; Gaps 14;
 370 TATTATTAGACAGGTTCTCACTCTCTCACTCCAGGCTGGAGTGCAGTGGGCACAAATCATA 429
 DB TTTTITTTGAGATGGATCTTACTCTCTCACTCCAGGCTGGAGTGCAGTGGGCACAAATCATA 429
 430 GCTCACTGTAACTCAAACTCTGGGCTCAAGTCACTCACTCACTCACTCACTCACTCACTCACT 489
 DB GCGGCTACAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3723
 490 GCTAGGACTACAGGACACACAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 549
 DB GCTGGGATGCGAGCATGGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3781
 550 ATGTATTATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 609
 DB GTAGAGACAGGTTTGGCATTTTGGCCAGGCTGGTCTGCACTCTGCACTCTGCACTCTGCACTCT 3841
 610 CCCAGGCTGGAGGAGTGGCATGTTGGCCAGGCTGGTCTGCACTCTGCACTCTGCACTCTGCACT 669
 DB CCGCCGCCACCTGACCTCCCAAGTGTGTTGTTATTAACAGGATGA-----GCCAGCAT 3894
 670 TGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 728
 DB ACCTAGCCCTGATTTTCAAGACAAATGAAACTGGAATTTAGATGTAATCTTTTCTTTTCTTTT 3954
 729 GGCTGATATTTTCTTGAACAGGATATCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 788
 DB TTTTITTTTITTTTITTTGAGAGGAGTCTCATGCTGCTCACTCCAGGCTGGAGTGGTG 4014
 789 GCGTAATAATAGCTCACTGCACTGCTCC--CTCTCTGGGCTCAAGCAATCCGCTGGCCTCAG 847
 DB GCGTGAATCTCGGCTCACTGCACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4014
 848 CATCTGAGTAGTGGGACTACAGGCTTGTGCCACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907

DB 4075 CCTCCCTAGTAGCTGGGACTACAGGCGTGTGCCACCACTCTCGGCTAATTTTGTATTT 4134
 QY 908 GATTTTGGTATAGAGAGGCTCTTGCTATGCTCAGGCTGATTTTATTTTATTTTATTTT 963
 DB 4135 CTAGTAGGAGCGGGTTTACCAGTTAGCCAGATGGTCTCTATTTTATTTTATTTTATTT 4194
 QY 964 ----AGACAAAGGCTCACTATGTTGCCATGATCCCCCACCCTCCACTCCOCAAAGTGCTC 1019
 DB 4195 TTTAAGACAGAACTCTGTTCTGCTACCTAAGGCTGGAGTGGTGTGATGTGCGCTCA 4254
 QY 1020 ATCTTATCTGTCATTTAGTTCAGTTCAGACAGATTTAGTTGTTTCCACTTTTTCACCAT 1079
 DB 4255 CTGCACTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4314
 QY 1080 ATGAATAATCTCCAGTGAATATTCATGTATACATTTGTGTGGGCATATGTTTTCATTTTC 1139
 DB 4315 TCTGTACTCTCATCCACTGAGTAGTTGGAATCACAGCGCTGGCCACCATTGCCCGCTA 4374
 QY 1140 TGTGGGTTTATATCTAGGAGTGAATTTGCTGGATCCCGGGTAAATTTTTCACAGGAGA 1199
 DB 4375 ATTTTGTGTATTTTATAGTAGATGGGTTTGTGCCACGTTGGCCAGGCTGCTGCTGAAC 4434
 QY 1200 GTTCAGGGGAGAAA-----ACTTGGGAAATGAAGCATGTTTGAATCA 1246
 DB 4435 TCCGCGCTCAAGTGAATTTCCCGCTTGGCTCCCAAGTGTGGGATACAGGGGTGA 4494
 QY 1247 GCAAGAGTGTGAGGGGTTTTCGGAGTTTATTTTATTTTATTTTGTGACAAATGTGCAGTT 1306
 DB 4495 GCCACATGCCAGCTGTTTATTTTATTTTATTTTATTTTAAAG--GCTGGGTATGCTAGCT 4552
 QY 1307 GATGAAGATCAAGTTATATCTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1366
 DB 4553 CATGCTCTGTAATCTTGAATCTGGAGAGCC--CGAGGAGGAGGAGTTCCTGAGACTAGGA 4611
 QY 1367 GAGTAAATCATGAAGCAGCTTTGAATACCAAAATTAAGGAGCTTGGCTGTAACAAATA 1456
 DB 4612 GTTCAAAACCAACTGCCCAATAGCAGGCTCTTTTAAATAAATAAATAAATAAATAAATAA 4671
 QY 1427 ATAAAAATCACAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1486
 DB 4672 TTTATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4731
 QY 1487 GGAGGAGTGGTGTGATCTCAGCTCACTGCACTTTCGCTCCCGGTTCAACAATTC 1546
 DB 4732 GGAGTGCAGTGGCGGATCTCAGCTCACTGCACTCCGCTCCGCTGCTGCTGCTGCTGCTGCT 4791
 QY 1547 TCCGTCTTCCGCTCCCAAGTGTGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 1605
 DB 4792 TCCGTCTTCCGCTCCCAAGTGTGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 4851
 QY 1606 --TTTTTGTATTTTATAGTAGATGGGATTTCACTTTTGTGGCCAAAGTGGTCTCAAACT 1663
 DB 4852 TTTTITTTGATTTTATAGTAGACAGGTTTCACTGTTTAAACAGGATGCTCTCAATCT 4911
 QY 1664 TTTTGTCTGTCATAATTTGTTGTAATTTGTTTCTTTTCTGCTGAGGTAGGCCCCCAGACCA 1723
 DB 4912 CCTGAC-----TTGCTGATCCACCCACTCAGCTCCCAAGTGTCTGGATATACAGCGCT 4966
 QY 1724 AAAAAATAATCTTAGAATCCAAATCAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1783
 DB 4967 GAGCCACAGCGCTGGCCCT-----GGCCTATCTTTTAAACCTTTTATTTTGGAGAA 5019
 QY 1784 CACAGTGTGACAGGCGCTCAGGAGTAGTGTGATCTCTGCTCGAAAGAGAAATAGATG 1843
 DB 5020 AAAAAATCAGAGGTGCCATTTGGCTTTTATGATGTCAAGTGAAGTTGAAAAAATTTT 5079
 QY 1844 AAAAAATTTCTCCGCGCAGGCGTGTGCTCATCTGTAATCCAGCACTTTTGGGAGGC 1903
 DB 5080 TTTTAAATGGGCTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5139
 QY 1904 CAAGGCATGTGGATCA-----CCTGAGGTGAGGAGT 1934

QY	1621	TAGAGATGGGATTTCACTTTGTTGGCCAAAGCTGGTCTCAAACTTTTTCGTGTCATAATTG	1680	PR	07-JUN-2000; 2000US-0209467.
Db	3805	TACAGATGGGTTTACCAATGTTGGTGGCTAGGCTGGTCTGAACTCTGACATCATGATCCG	3864	PR	28-JUN-2000; 2000US-0214886.
QY	1681	TTGTAACATATTTGTTCCCTTTTGTGAGGTAGGGCCCCAGACCAAAAAATAATCTTAG	1740	PR	30-JUN-2000; 2000US-0215135.
Db	3865	TCCCCCTCGACCTCCAAAGTGTAGGATTACAGGCATGAGCCACTGTGCCAGCC----	3920	PR	07-JUL-2000; 2000US-0216647.
QY	1741	AATCAAAATCAGTGTGTTGGTTGACCACTGTCACTTGAGAACCACTGTGACCAAGGC	1800	PR	2000US-0216880.
Db	3921	---GCTAATTTATTTTGTACAGACAGAGTTTCTCCATGTTGCCAGGCTGATCTGGAA	3977	PR	2000US-0217487.
QY	1801	CTCAGAGTAGAGGTATCTCTCTCGAAAGAGAAATAGAAATATTTCTCCGGGCC	1860	PR	2000US-0217496.
Db	3978	CTCCTGGGCGAAGTGAACCATCTGCCTCGGC---CTCCAAAGTCTGGGATTACAGGCC	4034	PR	2000US-0218290.
QY	1861	AGCGTGGTGGCTATGCTGTAAATCCAGCACTTTGGGAGGCCAAGGCATGTGGATCAC	1920	PR	2000US-0220964.
Db	4035	GGGTGGGTGGCTACGGCTGTAAATCCAGCACTTTGGGAGGCCAAGGCAGGTGGATCAG	4094	PR	2000US-0225214.
QY	1921	CTGAGTTCAGGAGTTCAAAACCAAGCTGGCCCAACATGTTGAAACCCGCTCTCTACTAAA	1980	PR	2000US-0225266.
Db	4095	CUGAGTCCGGAGTTCAAGACCGCTGGCCCAACAGGTGACCCCTGTCTCTACTAAA	4154	PR	2000US-0225267.
QY	1981	ATACAAAAATTAGCTAAAGTGTGGTGGCGCATGCTGTAAATCCAGCTACTTTGGGAGGGT	2040	PR	2000US-0225270.
Db	4155	ATAC-AAAAATTAGCCACGTGTGGTGGCGCATGCTGTAAATCCAGCTACTTTGGGAGGGT	4213	PR	2000US-0225447.
QY	2041	GAGCGAGGAGAAATTTCTTGAACCCGGGAGGAGAGGTGCGAGTGAAGGATGATCACCA	2100	PR	2000US-0225757.
Db	4214	GAGCGAGGAGAAATTTCTTGAACCCGGGAGGAGAGGTGCGAGTGAAGGATGATCACCA	4273	PR	2000US-0225758.
QY	2101	CTGCACTCCAGCTGGGGAGAGAGCGAGCTTCTCTCAAAAACAAACAAACAAAGA	2160	PR	2000US-0226279.
Db	4274	CTGCACTCCAGCTGGGGAGAGAGAGCTTCTCTCAAAAACAAACAAACAAAGA	4333	PR	2000US-0226681.
QY	2161	ATTAAGCAAAATTAGACA 2177		PR	2000US-0226686.
Db	4334	ACAAACAAACAAACAA 4350		PR	2000US-0227182.
RESULT 14				PR	2000US-0227009.
ABL98014				PR	2000US-0228924.
ID	ABL98014	standard; DNA; 13216 BP.		PR	2000US-0229287.
XX	AC	ABL98014;		PR	2000US-0229343.
XX	DT	21-JUN-2002 (first entry)		PR	2000US-0229344.
XX	DE	Human testicular antigen encoding DNA fragment SEQ ID NO: 2666.		PR	2000US-0229509.
XX	KW	Human; testicular antigen; testes; cancer; metastasis; immune disorder;		PR	2000US-0230437.
XX	KW	reproductive system disorder; urinary system disorder; gene therapy;		PR	2000US-0230438.
XX	KW	cardiovascular disorder; respiratory disorder; neurological disorder;		PR	2000US-0231242.
XX	KW	gastrointestinal disease; infection; cytostatic; gene; ds.		PR	2000US-0231243.
XX	OS	Homo sapiens.		PR	2000US-0231413.
XX	PN	WO200155317-A2.		PR	2000US-0231414.
XX	PD	02-AUG-2001.		PR	2000US-0232080.
XX	PF	17-JAN-2001; 2001WO-US01329.		PR	2000US-0232081.
XX	XX	31-JAN-2000; 2000US-0179065.		PR	2000US-0232397.
XX	PR	04-FEB-2000; 2000US-0180628.		PR	2000US-0232398.
XX	PR	24-FEB-2000; 2000US-0184664.		PR	2000US-0232400.
XX	PR	02-MAR-2000; 2000US-0186350.		PR	2000US-0232401.
XX	PR	16-MAR-2000; 2000US-0189874.		PR	2000US-0233063.
XX	PR	17-MAR-2000; 2000US-0190076.		PR	2000US-0233064.
XX	PR	18-APR-2000; 2000US-0198123.		PR	2000US-0233065.
XX	PR	19-MAY-2000; 2000US-0205515.		PR	2000US-0233065.

PR	20-OCT-2000;	2000US-0241785
PR	20-OCT-2000;	2000US-0241786
PR	20-OCT-2000;	2000US-0241787
PR	20-OCT-2000;	2000US-0241808
PR	20-OCT-2000;	2000US-0241809
PR	20-OCT-2000;	2000US-0241826
PR	01-NOV-2000;	2000US-0244637
PR	08-NOV-2000;	2000US-0246474
PR	08-NOV-2000;	2000US-0246475
PR	08-NOV-2000;	2000US-0246476
PR	08-NOV-2000;	2000US-0246477
PR	08-NOV-2000;	2000US-0246478
PR	08-NOV-2000;	2000US-0246523
PR	08-NOV-2000;	2000US-0246524
PR	08-NOV-2000;	2000US-0246525
PR	08-NOV-2000;	2000US-0246526
PR	08-NOV-2000;	2000US-0246527
PR	08-NOV-2000;	2000US-0246528
PR	08-NOV-2000;	2000US-0246532
PR	08-NOV-2000;	2000US-0246609
PR	08-NOV-2000;	2000US-0246610
PR	08-NOV-2000;	2000US-0246611
PR	17-NOV-2000;	2000US-0246613
PR	17-NOV-2000;	2000US-0249207
PR	17-NOV-2000;	2000US-0249208
PR	17-NOV-2000;	2000US-0249209
PR	17-NOV-2000;	2000US-0249210
PR	17-NOV-2000;	2000US-0249211
PR	17-NOV-2000;	2000US-0249212
PR	17-NOV-2000;	2000US-0249213
PR	17-NOV-2000;	2000US-0249214
PR	17-NOV-2000;	2000US-0249215
PR	17-NOV-2000;	2000US-0249216
PR	17-NOV-2000;	2000US-0249217
PR	17-NOV-2000;	2000US-0249218
PR	17-NOV-2000;	2000US-0249244
PR	17-NOV-2000;	2000US-0249245
PR	17-NOV-2000;	2000US-0249264
PR	17-NOV-2000;	2000US-0249265
PR	17-NOV-2000;	2000US-0249267
PR	17-NOV-2000;	2000US-0249299
PR	17-NOV-2000;	2000US-0249300
PR	01-DEC-2000;	2000US-0250160
PR	01-DEC-2000;	2000US-0250391
PR	05-DEC-2000;	2000US-0251030
PR	05-DEC-2000;	2000US-0251988
PR	05-DEC-2000;	2000US-0251989
PR	11-DEC-2000;	2000US-0251990
PR	05-JAN-2001;	2000US-0254097
XX		2001US-0259678

CC protein fragment of the invention.

Search completed: May 20, 2003, 14:53:49
Job time : 1583 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 14:24:42 ; Search time 134 Seconds
(without alignments)
9497.825 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150

Sequence: 1 aggaattcatcatttaaat.....tgagtttccagcctggccc 4150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	349.8	8.4	12394	4	US-09-488-856A-10
2	345	8.3	84495	4	US-09-797-906-3
3	321	7.7	35060	3	US-08-814-095-7
4	316.6	7.6	59065	4	US-09-813-817-3
5	316.6	7.6	59065	4	US-09-978-197-3
6	308.2	7.4	162450	4	US-09-345-882-1
7	307.6	7.4	43950	4	US-09-735-934A-3
8	307.2	7.4	70000	4	US-09-851-896-3
9	303	7.3	43950	4	US-09-735-934A-3
10	301.8	7.3	246240	2	US-08-724-394A-20
11	301.8	7.3	246240	2	US-08-724-394A-21
12	301.8	7.3	246240	2	US-08-724-394A-22
13	297.2	7.2	70000	4	US-09-851-896-3
14	296.6	7.1	59065	4	US-09-813-817-3
15	296.6	7.1	59065	4	US-09-978-197-3
16	288	6.9	21234	4	US-09-810-671-3
17	286.8	6.9	35060	3	US-08-814-095-7
18	285.8	6.9	162450	4	US-09-345-882-1
19	283.8	6.8	12394	4	US-09-488-856A-10
20	282.8	6.8	246240	2	US-08-724-394A-20
21	282.8	6.8	246240	2	US-08-724-394A-21
22	282.8	6.8	246240	2	US-08-724-394A-22
23	275	6.6	84495	4	US-09-797-906-3
24	270.4	6.5	72604	4	US-09-268-992-7
25	270.4	6.5	72604	4	US-09-657-474-7
26	267.2	6.4	45546	4	US-09-146-053-6
27	263	6.3	98844	4	US-09-791-211-10

C 28	262.8	6.3	29629	4	US-09-729-995-3	Sequence 3, Appli
C 29	261.2	6.3	17327	1	US-07-906-871-15	Sequence 15, Appli
C 30	260.6	6.3	7676	1	US-08-451-777A-7	Sequence 7, Appli
C 31	260.6	6.3	7676	2	US-08-451-778A-7	Sequence 7, Appli
C 32	260.6	6.3	7676	2	US-08-998-208-7	Sequence 7, Appli
C 33	260.6	6.3	7676	5	PCT-US95-06743-7	Sequence 7, Appli
C 34	259.4	6.3	17327	1	US-07-906-871-15	Sequence 15, Appli
C 35	258	6.2	152331	3	US-09-128-155-16	Sequence 16, Appli
C 36	258	6.2	152331	3	US-09-128-155-16	Sequence 16, Appli
C 37	258	6.2	176373	3	US-09-128-155-17	Sequence 17, Appli
C 38	257.2	6.2	29629	4	US-09-729-995-3	Sequence 3, Appli
C 39	256.8	6.2	4326	2	US-08-852-807-12	Sequence 12, Appli
C 40	256.8	6.2	13674	2	US-08-852-807-1	Sequence 1, Appli
C 41	254.8	6.1	176373	3	US-09-128-155-17	Sequence 17, Appli
C 42	248.8	6.0	99500	2	US-09-798-096-10	Sequence 10, Appli
C 43	244	5.9	4326	2	US-08-852-807-12	Sequence 12, Appli
C 44	244	5.9	13674	2	US-08-852-807-1	Sequence 1, Appli
C 45	243.4	5.9	99500	4	US-09-798-096-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-09-488-856A-10
; Sequence 10, Application US/09488856A
; Patent No. 6316259
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Robert McKay
; APPLICANT: Madeline M. Butler
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLYCOGEN SYNTHASE KINASE 3 ALPHA
; FILE REFERENCE: RFS-0115
; CURRENT APPLICATION NUMBER: US/09/488,856A
; CURRENT FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 10
; LENGTH: 12394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (115)...(397)
; NAME/KEY: CDS
; LOCATION: (2438)...(2625)
; NAME/KEY: CDS
; LOCATION: (5639)...(5722)
; NAME/KEY: CDS
; LOCATION: (5864)...(5974)
; NAME/KEY: CDS
; LOCATION: (7902)...(8032)
; NAME/KEY: CDS
; LOCATION: (8121)...(8227)
; NAME/KEY: CDS
; LOCATION: (9197)...(9294)
; NAME/KEY: CDS
; LOCATION: (9375)...(9470)
; NAME/KEY: CDS
; LOCATION: (9898)...(10084)
; NAME/KEY: CDS
; LOCATION: (10431)...(10523)
; NAME/KEY: CDS
; LOCATION: (11713)...(11786)
US-09-488-856A-10

Query Match 8.4%; Score 349.8; DB 4; Length 12394;
Best Local Similarity 55.2%; Pred. No. 9.6e-76;
Matches 1016; Conservative 0; Mismatches 742; Indels 81; Gaps 14;

OY 370 TATTATTAGACAGGTTCTCACTCTGTGTCACCCAGGCTGGAGTGGCAGTGCACATCATTA 429
Db 3604 TTTTITTTGATGAGTCTTCTCTGTGTCACCCAGGCTGGAGTGGCAGTGCATCTTG 3663

QY 430 GCTCACTGTAACCTCAAACTCTCGGGCTCAAGTGATCTACTACCTACCTCCAGAGTA 489
DB 3664 GCCCGGTACAACCTCTGCTCCAGGTTCAAGCGATCTTGCTGCTCAACCACTCAGTA 3723
QY 490 GCTAGGACATAGGACACACAGCCATACCTGCTCAATTTTCTTTTAAATTTCAATTT 549
DB 3724 GCTGGGATTCAGGCAATGGCCACATAGCTGCTCAATTTTCTTTTAAATTTTAA 3781
QY 550 ATGTAATTCATTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 609
DB 3782 GTAGAGACAGGGTTTGGCATGTTGGCCAGGCTGGTCTCGAATCTCTGACCTCAAGTAAT 3841
QY 610 CCCAGGCTGGAGGCGAGTGGCATGTTGGCCAGGCTGGTCTCGAATCTCTGACCTCAAGTAAT 669
DB 3842 CGCCCCCACTCTGAGCTCCCAAGGTTGGTGTATTACAGGCATGA-----GCCAGCAT 3894
QY 670 TGATCTCTCTGCTCAGCTCCCAAGTCTGGACTACAAACAGTGPCA-CCATGCTT 728
DB 3895 ACCTAGCCCTGATTTTCAAGACAACTGAAACTGGATTTAGATGTGAATCTTTTCTTTT 3954
QY 729 GCTGATATTTTCTTTTGAACAGGGTATCACTCTGTTGCCAGGCTGGAGTACAGTG 788
DB 3955 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 4014
QY 789 GGTATAATATAGCTCACTGCAAGCTCC-CCCTCTGGGCTCAAGCAATCGCTGGCTCAG 847
DB 4015 GCGTGATCTCGGGTCACTGCAACCTTCGCGCTGCGGGTTCAGCGGATTTCTCTGCTCAG 4074
QY 848 CATCTAGTAGTGGGACTCAGGCTTGTGCCACAGCCCGCCAGCTTAAGTTTAAAT 907
DB 4075 COTCCCTAGTAGTGGGACTCAGGCTGTGCCACACTCTCGGCTAAATTTTCTTTTCTTT 4134
QY 908 GATTTTGGTATAGAGAGGCTTCTGCTATGTTGCTCAGGCTGTATTTTATTTGTTG--- 963
DB 4135 CTAGTAGGAGCGGGTTTCAACGAGTTAGCCAGGATGCTCTATTTTCTTTTCTTTTCTTT 4194
QY 964 ---AGACAAGTCTCACTATGTTGCCATGATCCCCCACTCCACTTCCCAAGTGCTC 1019
DB 4195 TTTAAGACAGAACTCTGCTCTGCTCACTAAGGCTGGAGTGGTGTGATGCTCGGCTCA 4254
QY 1020 ATCTATCTGTTATCTAGTCACTAGTTCACAGACATTTAGTGTGTTTCCACTTTTTCACCAAT 1079
DB 4255 CTGCAACCTCTGCTCTGCTGCTTCAAGGCTGCAACCTCTGCTCTGCTGCTGCTGCTGCT 4314
QY 1080 ATGAATATATCTCAGTGAATATATATATATATATATATATATATATATATATATAT 1139
DB 4315 TCTTGTACCTCATCCACTGATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4374
QY 1140 TGTGGGTTTATATCTAGGAGTGAATTTGCTGATCCCGGTAATTTTTCACAGGCGA 1199
DB 4375 ATTTTGTATTTTATAGTAGAGTGGGTTTGGCCAGTGGCCAGGCTGGTCTCGAAC 4434
QY 1200 GTTCAGGGGAAGAAA-----ACTTGGGAAATGAAGCATGTTTGAATAATCA 1246
DB 4435 TCTTGGCTCAAGTGAATTTCTCTGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 4494
QY 1247 GCAAGTGCAGGGTCTTTTCTGAGTGTATTTATATATATATATATATATATATATATAT 1306
DB 4495 GCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4552
QY 1307 GATGAAGATACAAGTATATCTAGTGAAGTGAATTAAGCTGGAATAGGCGGTCA 1366
DB 4553 CATGCTGTATCTTGAACCTTGAGAGCC-CGAGGCGAGGAGGATGCTCGAGACTAGGA 4611
QY 1367 GAGTAAATCATGAAGCACTTTGAATACCAAAATTAAGGAGCTTGGCTGTAAACAAATA 1426
DB 4612 GTTCAAAACCAACCTGCCACATAGCCAGTCTTTTAAATATAATATAATATAAT 4671
QY 1427 ATAAAAATCAAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1486
DB 4672 TTTATCTATTTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4731

QY 1487 GGAGGCGAGTGTGTGATCTCAGCTCAGTCAACTTTTGGCTCCCGGTTTCAAGCAATTC 1546
DB 4732 GGAGTGCAGTGGGCGGATCTCAGCTCAGTCAAGCTCCCGCTCTCTGGTTTCAAGCAATTC 4791
QY 1547 TCTGCTTCAAGCTCCCAAGTGTAGTGGGACTCAGGCACTTCCACCATGCCCCAGCTGA- 1605
DB 4792 TCTGCTTCAAGCTCCGAGTGTAGTGGGACTCAGGCACTTCCACCATGCCCCAGCTGA- 4851
QY 1606 --TTTTTGTATTTTGTAGTAGAGGATTTCACTTTTGGCCAAAGCTGGTCTCAAACT 1663
DB 4852 TTTTGTATTTTGTATTTTGTAGTAGAGGATTTCACTTTTGGCCAAAGCTGGTCTCAAACT 4911
QY 1664 TTTTGTGTCAATAATTTTGTATTTTGTAGTAGAGGATTTCACTTTTGGCCAAAGCTGGTCTCAAACT 1723
DB 4912 CTTGAC-----TTCGTGATCCACCCACCTCAGCTCCCAAGGCTGGGATTCAGGCGT 4966
QY 1724 AAAAATAAATCTTTAGAAATCAAAATCACTGTGTGGTTTGGACCACTGTCTTGTAGAAC 1783
DB 4967 GAGCCACCAAGCTGCGCTTGGCTTTCATGTCAGCAATTAAGTTGAAAAAATTTT 5019
QY 1784 CACAGTGTGACAGGCGCTCAGGAGTGTAGAGTGTATCTGTCTGCGAAGAGAAATAGATG 1843
DB 5020 AAAAATCAGAGGTCGCTTGGCTTTCATGTCAGCAATTAAGTTGAAAAAATTTT 5079
QY 1844 AAAATATTTCCGGCCAGGCGTGGTGTATGCTGTCTGTAATCCCGACACTTTTGGAGGC 1903
DB 5080 TTTTAAAGTGGGCTGCTGGCGCGTGTCTCAGGCTGTATCCCGACACTTTTGGAGGC 5139
QY 1904 CAGGCAATGTGATCA-----CCTGAGGTCAGGAGT 1934
DB 5140 TGAGCGCTGTGATCATGAGTTCAGGAGGCTGAGGAGTGCATCAACAAGTTCAGGAGA 5199
QY 1935 TCAAAACAGGCTGCGCAACATGTTGAACCCCGTCTCTACTATAAATACAAAAATTAG 1994
DB 5200 TCGACACATCTCTGCTAAGTGTGGTGAACCCCATCTCTACTATAAATACAAAAATTAG 5258
QY 1995 CTAAGTGTGTGGCGCATGCTGTATCCAGCTACTTTGGGAGGCTGAGGAGGAGATTT 2054
DB 5259 CTGGCGTGTGTGCTGATGCTGTATCCAGCTACTTTGGAAGGCTGAGGAGGAGATTT 5318
QY 2055 TCTTGAACCCGGGAGGCGAGGTTGCTGTAAGCGAGATACACACCTGCTCAGCTT 2114
DB 5319 GCTTGAACCCGGGAGGCGAGGTTGCTGTAAGCGAGATACACACCTGCTCAGCTT 5378
QY 2115 GGGGAGAGAGCGAGACTTCTCTCAAAAAACAAAAA 2153
DB 5379 -GGCAACAGACAGACTCTCTCAAAAAA 5416

RESULT 2

US-09-797-906-3/c
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zilanghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)....(84495)
; OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3

Query Match 8.3%; Score 345; DB 4; Length 84495;
Best Local Similarity 55.0%; Pred. No. 3.7e-74;
Matches 979; Conservative 0; Mismatches 715; Indels 85; Gaps 12;

QY 445 AAACCTCCTGGGCTCAAGTATCCTACTACCTCAGCC-TCCAGAGTAGTAGGACTCAGG 503
DB 18461 AACTCTCGGGCTGAGCCATCCACCTGCTGGCCCTCCAAAGTGGGGGATTACAAG 18402

QY 504 CACACACAGCCATACCTGGCTAAATTTTTTTTTTTTAAATTTTCATTTTATGATTCATTTTC 563
DB 18401 CATTAAGTCAACAGCCCTGGCATTTTATATATATATATATATATATATATATATATTTT 18342

QY 564 TTTCTTTTGTG-TTGTTGTTGTTTTCAGATAGGCTCTCACTTTGTTACCCAGGCTGGAG 622
DB 18341 TT 18342

QY 623 CGAGTGGCATGTGACAGCTGA--GCAGCCTTGCATCTCCTGGCTCAAGTGATCCCTCCNG 680
DB 18281 GCAGTGGCATGATCTCAGCTCACTGCAACTTGGACTCTGGCTCCGGTTCACCAATTCCTCG 18222

QY 681 CCTCAGCCTCCCAAGTAGCTGGGACTACAAACACGCTGACCATGCTGGCTGATATTTT 740
DB 18221 CCTCAGCCTCTTAAGTAGCTGGGATTTACAGGCATGTGACCGCGCCAGCTAAATTTT 18162

QY 741 TTTTCTTGAAACAGGGTATCACTCTGTGCTGCCAGGCTGGAGTACAGTGGCGTAATAATAG 800
DB 18161 TTTTCTT-----TTTTTGTTTTTTTTTTTTTTTTGTAGAGCGGAGTTTCACTATATTAG 18109

QY 801 CTCACCTGACGCTCCCTCCTGGCTCAAGCAATCGCTGGCTCAGCATCTCTGAGTAGC 860
DB 18108 CCAGCGCGATCTCGAACTCCTGACCTCAAGTATCTGCGCCGCTCAGCTCCCAAGTGT 18049

QY 861 TGGGACTACAGCTTGTGCCACAGCCAGCCAGCTAAGTTTTTAAAAAATGATTTTGGTATA 920
DB 18048 TAAGATTACAGGCATGAGCCACCATCGCGGCTCTCTATTTTATAGTTTTTAAGACACAA 17989

QY 921 GAGGAGGTCTTCTATGTTGCTCAGGCTGATTTTTTATTTTGTGACACAGGCTCACTAT 980
DB 17988 GA----GTCCCACTCTGTGCGCCAGCTGGAGTGAGTGAGGCAATCATAGTCTACTGCA 17933

QY 981 GTTGCATATGATCCCCCACTCCACTCCCAAGTGTCTCATCTTATCTG--TTCAATTAGT 1038
DB 17932 GCTTTAAACTCTGGGCTAAAGCAATCTCTACCTCAGCTCCCGAGTAGCTACATGAC 17873

QY 1039 CAGTTGACAGCATTTAGTGTGTTTCCATTTTGGACATATGATAATACTCCAGTGA 1098
DB 17872 ATGCCACCATACCCAGCTAAATTTTTTAAATGTTTTTAAAGATGGGGTCTCACTATGTG 17813

QY 1099 ATATTCATGTATACATTTGTGTTGGCATATGTTTTCATTTCTGTTGGTTTATATCTAGG 1158
DB 17812 CCCAGGCTGTGGAATCTCTGGCTTAAGCGATCTCCCATCTTGGCTTTCCAAAGTG 17753

QY 1159 AGTGAATGTGTTATCCCGGTAAATTTTTCAGAGGAGAGTTCAGGGGAGAAATACT 1218
DB 17752 CTGGCATTTACAAGTGTGAGCCACCATGCTGGCCCAATATGATTTTTCAGAGTCACTT 17693

QY 1219 TGGGAAATGAGCATGTTTGAATATCAGAGAGTGGAGGGTTTTTCGGAGTTTAT 1278
DB 17692 GACTGCGATGG--AGAATAAATGGAGGAGTAAATAGTTTAAATAGGAGTATTTTA 17635

QY 1279 TTATATCTGTTGACAAATGTGCAATTTGATGAGATACAGTATATCTAAGTGAAGT 1338
DB 17634 ATGACCATACACAGCCCATCTCACTTATGAAATAATCATAAATTTACTGATCAATTTCT 17575

QY 1339 GAGAAATTAAGGCTGGAATAGGCGTTCAGATGAAATCATGAACACTTTGA----ATAC 1394
DB 17574 CTATTTTATTTGACATTTTAGACTATTTCCAGTTCTTTCTAATAATTAATAAGGACAC 17515

QY 1395 CAAATAAAGAGCTGCTGCTTAACAAATAATAAAAAATCAAAATTTTTTTTTTTTTT 1454
DB 17514 TGTAAATCAATCTTGAACCTAAATCTTTGGCAGCATCTCTAATTAATTTTCTTTCTTTT 17455

QY 1455 TTGAGAAAGAGCTTTGCTCTTTTCCCTCTGCTGGGAGGAGTGGTGTGATCTCAGCTAC 1514

DB 17454 TTTTGACAGAGTCTCAGCTGTTGCCAGGCTCGAGTGCAGTGTGTCAGTCAATCCAGCTCAC 17395

QY 1515 TGAACATTTGCGCTCCCGGTTCAAGCAATTCCTCTCAGCTCCCAAGTAGCTGGG 1574

DB 17394 TGAACCTCCACCTCCCTGGGTTCAAGTGAATTCCTCTCAGCTCCCAAGTAGCTGGG 17335

QY 1575 ACTACAGGCACTTCCACCATGCCAGCTGATTTTGTATTTTGTAGTAGAGTGGGATTT 1634

DB 17334 ATTACAGTGGGCGCCACCATGCTGCTAATTTTGTACTTTTAGCAGACAGAGGTTT 17275

QY 1635 C-ACCTTTGTTGGCCAAAGCTGGTCTCAAACTTTT----- 1666

DB 17274 CGCAATTTTGGCCAGGCTGGTCTCAAACTCTGCGATTTACAGGCGTGAGCCA 17215

QY 1667 -----TGCTCTCAATTTGTTGAATTTGTTTCCCTTTTGTGCTAGGTTAGGCGCCCCCA 1718

DB 17214 CCAGCGCCGACTTAATTTTCTAAATATAGATTTCTAGAGTGAATTTAAGTAAATAA 17155

QY 1719 GACCAAAAAATAAATCTTAGAATCCAAATCAGTGTGTTGTTGACCACTGCTCACTTG 1778

DB 17254 GTTAAAGGTATGAAATTTTGGCTATTTTAAATAATTTGTTTTTAAAGATFAAGATT 17095

QY 1779 AGAACACAGTGTGACAGGCGCTCAGAGTAGAGTGTATCTCTGCTCGAAAGAGAAATA 1838

DB 17094 TTGATCATTAATAATAAAAAATCTACTTTGCAAAATGGATATCTTTTCACTATCCAGTC 17035

QY 1839 GAAT-----GAAATATTTCCGGGCGCAGGCGGTGGTGGCTC 1874

DB 17034 AAATTATATCAGGGGAAATGTGCTTTTAAAGATGAAAACTAGGCCAGCGAGTGGCTG 16975

QY 1875 ATGCTGTATATCCAGCACTTTGGGAGGCCAAGGCATGTGGATCACTGAGGTCAGGAGT 1934

DB 16974 AGCGCTGTATATCCAGCACTTAGGAGGCTGAGTGGCGGATTTAC--AAGTCAAGAGT 16917

QY 1935 TCAAAACAGCCTGCGCAACATGTTGAACCCCGTCTCTACTAAAAATACAAAAATTTAG 1994

DB 16916 TCGAGACCATCTGCTGAACATGTTGAACCCCATCTCTACTAAAAACTAAAAAAATAG 16857

QY 1595 TTAAGTGTGTCGCGCATGCTGTAATCCAGCTACTTGGGAGGTTGAGGCGAGGAAAT 2054

DB 16856 CCGGCTGTCTCGGCGGCGCTGTAGTCCAGCTACTCGGAGGCTGAGGCGAGGAAAT 16797

QY 2055 TCTTGAACCCGGGAGCGAGAGTTGCAGTGAAGCGAGATCACACCACTCACTCCAGCT 2114

DB 16796 TCTTGAACCCGGGAGATGAGGTTGCAGTGAAGCGAGATTGACCACTCACTCCAGCT 16737

QY 2115 GGGGAGAGAGGAGACTTCTCTCAAAAAACAAAAA 2153

DB 16736 GGGCAACAGCGCCAGACTCTGATTTCAAAAAAGAAAAA 16698

RESULT 3

US-08-814-095-7
; Sequence 7, Application US/08814095
; Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Sored, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KOHN & ASSOCIATES
; STREET: 30500 No. 6025183thwestern Highway, Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible


```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 2391.00066
; FILING DATE: US/08/814,095
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Montgomery, Ilene N.
; REGISTRATION NUMBER: 38,972
; REFERENCE/DOCKET NUMBER: 2391.00066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (248) 539-5050
; TELEFAX: (248) 539-5055
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Cosmid including ACHE
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 7q22
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 4089..22464
; OTHER INFORMATION: /function= "ACHE Promotor"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 22465..22537
; OTHER INFORMATION: /function= "non-translated"
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 24090..25177
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "(translation start:
; OTHER INFORMATION: 24110)"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25524..26009
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27005..27274
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 27255..28007
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 27385..27387
; FEATURE:
;
; NAME/KEY: exon
; LOCATION: 28008..28129
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ACHE"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 28129..28131
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (34528..34895)
; OTHER INFORMATION: /function= "arsenite resistance
; OTHER INFORMATION: /gene=
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (34092..34958)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33779..33963)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33493..33591)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (33297..33408)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 5
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (32959..33094)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (32569..32628)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 7
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (32386..32468)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 8
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (31894..32080)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 9
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (31363..31534)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 10
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (31131..31284)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 11
; FEATURE:
; NAME/KEY: exon
; LOCATION: complement (30816..31011)
; OTHER INFORMATION: /gene= "AR"
; OTHER INFORMATION: /number= 12
; FEATURE:
; NAME/KEY: exon

```

```

, , LOCATION: complement (30470..30626)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 13
FEATURE:
, , NAME/KEY: exon
, , LOCATION: complement (30187..30274)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 14
FEATURE:
, , NAME/KEY: exon
, , LOCATION: complement (29945..30073)
, , OTHER INFORMATION: /gene="AR"
, , OTHER INFORMATION: /number= 15
FEATURE:
, , NAME/KEY: exon
, , LOCATION: complement (29664..29856)
, , OTHER INFORMATION: /gene="ARS"
, , OTHER INFORMATION: /number= 16
US-08-814-095-7

```

Query Match	7.7%	Score 321;	DB 3;	Length 35060;
Best Local Similarity	68.8%;	Pred. No. 1.8e-68;		
Matches 511;	Conservative 0;	Mismatches 205;	Indels 27;	Gaps
QY	1445	TTTTTTTTTTTGACAAAGAGCTTGTCTTTTACCCCTGGCTCGAGGGCAGTGGTGAT	1504	
DB	9802	TTTTTTTTTTTGACGAGAGTCTGCTCTATCGCCAGGCTGAGTGCAGTGGCAAT	9861	
QY	1505	CTCAGCTCACTGCAACTTTCGCTCCCGGGTTCGAAGCAATTCCTCGCTTCAGCTCCCA	1564	
DB	9862	CTCGGCTCACTGCAACTCTGCCACCCAGGTTCGAAGCAATTTTCCTGTGACGCTCCCA	9921	
QY	1565	AGTAGCTGGGACTCAGGCACATCCCAACCATGCCAGCTGATTTTGTATTTTAGTAGA	1624	
DB	9922	AATAGCTGGGATTACCACTGCCACACATGCCAGCTAATTTTATATTTTCAGTAGA	9981	
QY	1625	GATGGATTTCACTTTGTGGCCAAAGCTGTCTCAAACTTTTGTGTCTATAATTGTGT	1684	
DB	9982	GACAGGGTTTCACCATGTTGGCCAGGCTGTCTCAAAATCCTCGCCTCAAGTCATCTGCC	10041	
QY	1685	AAC-----TATTGTTCCTTTTGTGAGGTAGGGCCCCCAGACCAAAAAATA	1732	
DB	10042	CACCTCAGCCTCCAAAGTGTGGGATTACAGATGTGAGCCACTCACTCAGCCAGAAATC	10101	
QY	1733	AATCTTAGAATCCAAATCAGTGTGTGGTTTGACCACCTGTCACCTTGAGAAACCA	1785	
DB	10102	CCTCCCAATTAAAAATCAACAGAAACCCCAAGAAATCACATTTGTTGGCATTTGCTCT	10161	
QY	1786	-----CAGTGTGACAGGGCCTCAGGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAG	1839	
DB	10162	AGTAATATGTCATCTTGGGTGCAAGCGTCTAATGTCTAAACCTGCATCTGCATTTTA	10221	
QY	1840	AATGAAAATATCTCCGGGCCAGCGGTGGTGCCTATGCTGTAAATCCAGCACTTTGGG	1899	
DB	10222	CATTTTAAAGATGCTCCGGCCAGCGGTGGTGCCTCAGCCTGTAAATCCAGCACTTTGGG	10281	
QY	1900	AGGCCAAGGATGTGGATCACTGAGTCAAGGATTCAAAACAGCTGGCCCAACATGCT	1959	
DB	10282	AAGCCGAGGCAAGTGGATCACTTCAGGTCAAGAGTTCGAGACTTAGCTCAACCAATAGT	10341	
QY	1960	GAACCCCGTCTCTACTATAAAATACAAAAATATTAGTAGTGTGGGCGCATGCTGTA	2019	
DB	10342	GAACCCCATCTCTACTAAAAA-ATACAAAAATTAGCCAGGTGTGGTGGCGCATGCTGTA	10400	
QY	2020	ATCCAGCTACTTGGGAGGTGTAGGCAGGAGAAATTTCTTGAACCCGGGAGGAGGTGTG	2079	
DB	10401	ATCCAGCTACTTGAAGGAGGTGTAGGCAGGAGAAATGCTTTGAACCCGGGAGGAGAGTGTG	10460	
QY	2080	CAGTGAAGCGAGATCACACCACTGCACCTCCAGCTGGGGAGA-GAGCGAGACTTCCCTCT	2138	
DB	10461	TAGTGCATGTGATGTGCACATGCGCATCTCCATCCTGGGCAACAGGAGAAAACTCCGTCT	10520	
QY	2139	CAAAAAACAAAAACAAAAAGAA	2161	

```

||||| ||||||| ||| |||
Db 10521 CAACACAAACAAACAAACAGAAA 10543

RESULT 4
US-09-813-817-3/C
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01178
; CURRENT APPLICATION NUMBER: US/09/813,817
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-813-817-3

Query Match          7.6%; Score 316.6; DB 4; Length 59065;
Best Local Similarity 66.7%;   Fred. No. 2.8e-67;
Matches 561; Conservative    0; Mismatches 264; Indels 16; Gaps 16;

```

Db 11808 AATTAGCTGGTATGGGGCGGGTGGCT-TAAATCCAGCTACTTGGAGGCTGAGGCGG 11750
QY 2049 AGAATTTCTTGAACCCGGGAGGAGAGGTTGAGTGAAGCGAGATCACACCACTGCACTC 2108
Db 11749 AGAATTTCTTGAACCCGGGAGGAGAGTTCAGTGAAGCGGAAATTTGGCCACTGCACTC 11690
QY 2109 CAGCC-TGGGGGAGAGCGAGACTTCTCTCAAAAAACAAACAAACAAAGATTAAGC 2167
Db 11689 CAGCC-TGGGGGAGAGCGAGACTTCTCTCAAAAAACAAACAAACAAACAAATC 11630
QY 2168 A 2168
Db 11629 A 11629

RESULT 5
US-09-978-197-3/c
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001178DIV
; CURRENT APPLICATION NUMBER: US/09/978,197
; PRIOR FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: Human
US-09-978-197-3

Query Match 7.6%; Score 316.6; DB 4; Length 59065;
Best Local Similarity 66.7%; Pred. No. 2.8e-67;
Matches 561; Conservative 0; Mismatches 264; Indels 16; Gaps 7;
QY 1339 GAGAATTAAGCGTGAATAGGGCGTTTACAGAGTAAATATCATAGAGCACTTTTGAATACCAA 1398
Db 12464 GAGAGTGAGGCTGAAGTGCAGGAGCAGGAGGTTGTGGAGATGTGGAGAGAGGCGCACAT 12405
QY 1399 ATTAAGSAGCTTGGCTGTAAACAAATAATAAATAATCAATTTTTTTTTTTTTTTTGA 1458
Db 12404 TACGGAGGCTCTACTGTAGGCGCAATAA-AGGAATTTTTTTTTTTTTTTTTTTTGA 12346
QY 1459 GAAAGAGCTTGTCTTTCACCTGGCTGGAGGCGAGTGTGATCTCAGCTCACTGCA 1518
Db 12345 GATGAGTGTCTACTGTGTACCA-GGCTGAGTGCAGTGGCGTGTGCTGCTGCA 12287
QY 1519 ACCTTCGCTCCCGGTTCAAGCAATTCCTGCTTCCAGCTCCCAAGTAGCTGGGACTA 1578
Db 12286 ACCTTCGCTCCCGGTTCAAGTGTATCTCCTGCTCAGCTCCCAAGTAGCTGGGACTA 12227
QY 1579 CAGGCACTTCCCACTGCTCAATTTTGTATTTTGTAGAGATGGGATTTCACT 1638
Db 12226 CAGGCACTTCCCACTGCTCAATTTTGTATTTTGTAGAGATGGGATTTCACT 1638
QY 1639 TTGTTGGCAAGCTGCTCAACTTTTGTGCTGCTATAATTTTGTGTAAT 1688
Db 12166 ATGTTGGCAGGATGCTCAATCTTGTACTTTCATGATCTGCTGCTGCTGCTCCCA 12107
QY 1689 ATGTTGCTTTTGTGCTGAGTGGGCGGCGGAGCAACCAAAATAATCTTGAATCCAA 1748
Db 12106 AGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12047
QY 1749 TCAGCTGTGTTGTTTACCACCTCTCACTTTGAGAACACACAGTGTGACCGGCTCAGGAG 1808
Db 12046 TCCAGAACACTGGGAAGTTGTGGAGATATACAAAGTAGGAGAGACACACAGAGTGACA 11987

QY 1809 TAGAGTGATCTCTGTCTCGAAGAGAGAAATAGAAATGAAATATTTCTCCGGGCGGCGGTGG 1868
Db 11986 CAGTCAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAGATCA-CTCCTTCCAGCGGTCAG 11928
QY 1869 TGGCTCATGCTGTAAATCCAGCACTTTGGAGGCCAAGCACTGTGATCACCCTGAGGTC 1928
Db 11927 TGGCTCATGCTGTAAATCCAGCACTTTGGAGGCCAAGCACTGTGATCACCCTGAGGTC 11868
QY 1929 AGAGTTCAAAACAGCCTGCGCAACATGTTGAACCCCGCTCTCTACTAAATAATACAAA 1988
Db 11867 AGAGTTTCGAGACTAGCTTGGCCCAACATGCGCAACCCCATCTCTACTAAATAAT-ATAA 11809
QY 1989 AATTAGCTAAGTGTGGGCGCATGCTGTAAATCCAGCTACTTGGAGGCTGAGGCGG 2048
Db 11808 AATTAGCTGGGTATGGGGGCGGCTGTAAATCCAGCTACTTGGAGGCTGAGGCGG 11750
QY 2049 AGAATTTCTTGAACCCGGGAGGAGGTTTGCAGTGAAGCGAGATCACACCACTGCACTC 2108
Db 11749 AGAATTTCTTGAACCCGGGAGGAGGTTTGCAGTGAAGCGGAAATTTGTCACCTGCACTC 11690
QY 2109 CAGCC-TGGGGGAGAGCGAGACTTCTCTCAAAAAACAAACAAACAAAGATTAAGC 2167
Db 11689 CAGCC-TGGGGGAGAGCGAGACTTCTCTCAAAAAACAAACAAACAAACAAATC 11630
QY 2168 A 2168
Db 11629 A 11629

RESULT 6
US-09-345-882-1/c
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72794
; OTHER INFORMATION: 5-124-273 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88073
; OTHER INFORMATION: 5-127-261 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93714
; OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97122
; OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele

```
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
```

```

;
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108084..108130
; OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
;
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108127..108177
; OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
;
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108127..108177
; OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
;
; FEATURE:

```

Query Match	7.43;	Score 308.2;	DB 4;	Length 162450;
Best Local Similarity	66.7%;	Pred. No.5.3e-63;		
Matches 529;	Conservative	0;	Mismatches 233;	Indels 31;
				Gaps 5;

QY	1421	AAATAATAAATAATCAAAATTTTTTTTTTTTTTTTTTTTTTTGAGAAAGAGTCTTGCTCTTTTCACC	1480
Db	86851	AATGTAACAACAAGTGAAGAAATATTTTTTTTTTTTTTTTTTTGAGCGAGCTAGACTCTGT	-TGC
QY	1481	CTGGCTGGAGGACGTGGTGTCTCAGCTCACTGCAACTTTCGGCTCCCGSGTTCAG	1540
Db	86792	CGGGCTGGAGTGCAGTGGCCACCCTTGGCTCACICGACCTCCGGCTCTGGGTTCAG	867933
QY	1541	CAATTCTCTGTTCTCAGCCTCCCAAGTAGCTGGGACTACAGGCACCTCCCAACATGCCCA	1600
Db	86732	CAATTCTCTGCTCAGCCTCCTCAGTAGCTGGGATTTACAGGTGCCTGGCACCAAGCCCA	86673
QY	1601	GCTGATTTTGTGTTATTTTGTAGAGATGGGATTTCACTTGTGTGGCCAGAGCTGGTCTCAA	1660
Db	86672	GCTAATTTTGTATTTTGTGAGAGAGGGTTTACCATTGTAGCCAGATAGTCTCGA	866113
QY	1661	ACT-----TTTGTGCTGCTCAATTTGTTGTAACATTATGTTCTTTTCTGAGGT	1708
Db	86612	TCTCTGACCTCGTGATCCGCCACCTCGCTCCCAAAGTGCTGGGATTATAGGCGTGAG	86553
QY	1709	AGGGCCCCAGACCAAAAATAAATCTTTAGAAATCCAAATCAGTGTGTGGTTTGACCA	1768
Db	86552	CCACCCCACTTGGCCAAAAAGTAAAAACATTATATATAATTAAGCTTTTAAAGTGA	864933
QY	1769	CTGTCACTTGAAGACCAAGT-----TGACCAGGGCCTCAGAGTAGAGTGATCTC	1821
Db	86492	CATGGGGAAATCCACCCCTTGATTTCAATGATCATGTCAAATAACTTAAATAATGTTCCC	864333
QY	1822	TGCTCGAAGAGAAATAGAATGAATAATTTCTCCGGGCCAG-----CGTGGTGG	1871
Db	86432	TGGCAGAACCAAGCAATCTTCAAAGCTTGCTATTGAATAGAAGTACTGTACTTTGGCT	86373
QY	1872	CTCATGCTGTAAATCCAGCACTTTGGGAGGCCAAGGCATGTGGATCACTCGAGTCAAG	1931
Db	86372	GGAATATTTGTACTCCAGCACTTTGGGAGGCCAAGGCAGTGGATCACTCAGTCAAG	863113
QY	1932	AGTTCAAACACCAAGCTGGCCAAACATGCTGAACCCCGTCTCTACTAAAAATACAAAAAT	1991
Db	86312	AGTTCCGACCCAGCCTGCCAAATGTTGAACCTCGTCTCTACTAAAAATAC-----	86254
QY	1992	TAGCTAAGTGTGGTGGCGCATGCTCTGAATCCAGCTACTTGGGAGGGTGGAGCAGAGA	2051
Db	86253	TAGCCAGGCATGGTGGCGGGTGCCTGTAAATCCAGTACTCTGGGGCTGGAGCAGAGA	86194
QY	2052	ATTTCCTGAACCCGGAGGAGAGTTCAGTGAAGCGAGATCACACCACTGCATCCAG	2111
Db	86193	ATTGCTGAACCCGGAGGCGGAGGTTGCGAGTGACCGGATATCGACCACTGCATCCAG	86134
QY	2112	CCTGGGGAGAGCGGAGACTTCCCTCAAAAAAACAACAAAGAAATTAAGCAAT	2171
Db	86133	CCTGGGCGACAGGTGAGACTCCCTCTCAAAAAAACAACAAAGAAATTAAGCAAT	86074
QY	2172	TAGACATTTGCAGA	2184
Db	86073	TGGCAGCTTAAAGA	86061

[illegible]

Db 5595 TGAGGAGGAGATTGCTTGAACCCGGGAGGAGGAGGTTGCCGTGAGCCAGATCATGCC 5536
QY 2100 ACTGCATCCAGCTCGGGGAGGAGGAGGACTTCCCTCAAAAAAACAACAAAAG 2159
Db 5535 ACGGCACTCCAGCTCGGGTGACAGAGTAAGATTCTGTCTAAAAAAAAGAAAAA 5476
QY 2160 A 2160
Db 5475 A 5475

RESULT 10
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 7.3%; Score 301.8; DB 2; Length 246240;
Best Local Similarity 68.0%; Pred. No. 2.4e-63;
Matches 493; Conservative 0; Mismatches 177; Indels 55; Gaps 3;

QY 1441 TTTTITTTTTTTTGAGAAAGAGTCTTCTCTTCCCTTGGTGGGAGGAGTGGTG 1500
Db 2626 TTAITTTTATTTTIGAGAGTGTCTACCTGTGCGCCAGGCTGGATGCGAGTGGIG 2685
QY 1501 TGATCTCAGCTCAGTGAATTTTGGCTCCCGGGTTCAACAATCTCTGCTTCAGCCT 1560
Db 2686 CAGTCTTGGTCTAGTCAATCCCGAGTCCCGAGGTTCAACAGATTCTCTGCTCAGTCT 2745

QY 1561 CCCAAGTAGCTGGAGCTACAGGCATTTCCACCATGCCAGCTGATTTTGTATTTAG 1620
Db 2746 CCCAAGCAGCTGGGATTACAGTGCCTGCCACCATGCCAGCTAATTTTGTATTTAG 2805
QY 1621 TAGAGTAGGAGTTTACATTTTGTGGCCCAAGCTGTCTCAAACTTTTGTCTGCTCAATTTG 1680
Db 2806 TAGAGCAGGGTTTCGCCCTGTGTGGCCAGGCTGTCTCGAACT----- 2848
QY 1681 TTGTAACTATTGTTCCCTTTTCTGAGTAGGGCCCCCAGACCAAAAAATAAATCTTAG 1740
Db 2849 -----CTGACCTCAAGTGAATCTACCCGCC 2873
QY 1741 AATCCAAATCAGTGTGTGTTTGACCACTGTCTACTTTGAGAACCAACAGTGTGACCGAGGC 1800
Db 2874 TCGGCTCTCCCAAGTGTGGGATTACAGGTGTGAGCCACTGCGCCAGCGGGGTTCATC 2933
QY 1801 CTCAGGAGTAGAGTGTCTCTCTCGAAAGAGAAATAGATGAATATTTCTCCGGGCC 1860
Db 2934 CTTAATA-CATACATTAGAGATATAGATTCTGTTTTTATCTAAAGAGTCTTTATAAGGCC 2992
QY 1861 AGGCGTGTGCTCATCGCTGTATCCAGCACATTTGGGAGGCCAAGGCATGTGGATCAC 1920
Db 2993 GGGCGGCTGTGCTCACGCTGTATCCAGCACATTTGGGAGGCCGAGCGGGCGGATCAC 3052
QY 1921 CTGAGTCTCAGAGTTCAAAACCCAGCCCTGGCCAAACATGTGAACCCCGTCTCTACTAAAA 1980
Db 3053 --GAGTCTCAGAGATCGAGACCATCCCGGTAAACCGGTGAACCCCGTCTCTACTAAAA 3110
QY 1981 ATACAAAAAATTAGTAAAGTGTGGTGGCGCATGCTCTAATCCAGCTACTTGGAGGGT 2040
Db 3111 ATACAAAAAATTAGCGGGCGTAGTGGGGCGCTCTAGTCCAGCTACTTGGAGGGT 3170
QY 2041 GAGCAGGAGAAATTTCTTGAACCCGGGAGGAGGAGTTCAGTGAAGCGAGATCACACCA 2100
Db 3171 GAGCAGGAGAAATGGCGTGAACCCGGGAGGAGGAGTTCAGTGAAGCGAGATCACACCA 3230
QY 2101 CTGCACTCCAGCTGGGGGAGAGGAGAGTCTCTCTCAAAAAACAACAAACAAAGA 2160
Db 3231 CTGCACTCCAGCTGGGGGAGAGGAGAGTCTCTCTCAAAAAACAACAAACAAAGA 3290
QY 2161 ATTAA 2165
Db 3291 AAAAA 3295

RESULT 11
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30


```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
;
US-08-724-394A-21

Query Match          7.3%; Score 301.8; DB 2; Length 246240;
Best Local Similarity 68.0%; Pred. No. 2.4e-63;
Matches 493; Conservative 0; Mismatches 177; Indels 55; Gaps 3;

QY 1441 TTTT...TTTGTGAGAAAGACTTGTCTTTTCCACCTGGCTGGAGGCGAGTGGTG 1500
      |||...|||
DB 2626 TTATTTTATTTTGTAGAGCTTGTCTACCCCTGCGCCAGGCTGGAGTGCAGTGGTG 2685

QY 1501 TGAATCTAGCTCAGTCAACTTTCGCCCTCCCGGGTTCGAAGCAATTCCTCGCTTCAGCCT 1560
      |||...|||
DB 2686 CAGTCTTGCTCAGTCAATCCCAAGTTCGAAGCAATTCCTCGCTTCAGCCT 2745

QY 1561 CCCAAGTAGCTGGGACTCAGGCACTTCCACCATGCCAGTGAATTTTGTATTTTATAG 1620
      |||...|||
DB 2746 CCCAAGTAGCTGGGACTCAGGCACTTCCACCATGCCAGTGAATTTTGTATTTTATAG 2805

QY 1621 TAGAGATGGGATTTACATTTTGTGCGCAAGCTGGTCTTCAAACTTTTGTCTGTCTATAATTG 1680
      |||...|||
DB 2806 TAGAGATGGGATTTACATTTTGTGCGCAAGCTGGTCTTCAAACTTTTGTCTGTCTATAATTG 2848

QY 1681 TTGTAACATTTGTCCTTTTGTGAGGTAGGGCCCCCAGACCAAAATAAATCTTAG 1740
      |||...|||
DB 2849 -----CCTGAGCTCAAGTGAATCTACCCGCC 2873

QY 1741 AATCCAAATCAGTGTGTTGTTGACCACTGTCTTTCAGAACCACTGTGACCAAGGCG 1800
      |||...|||
DB 2874 TCGGCTCCCAAGTGTGGATACAGTGTGAGCCACTGCGCCCGCCAGCGGGGTTCATC 2933

QY 1801 CTCAGGAGTAGAGTGAATCTCTGCTCGAAGAGAAATAGAATGAAATATTTCTCCGGGCC 1860
      |||...|||
DB 2934 CTTAATA-CATACATTAGAGATATAGATCTCTTTTATCTAATAAGTCTTTATAAGGCC 2992

QY 1861 AGGCGTGTGCTATGCTGTATCCAGCACTTTGGGAGCCCAAGGCATGTGGATCAC 1920
      |||...|||
DB 2993 GGGCGCGTGTGCTACAGCTGTATCCAGCACTTTTGGGAGCCCGAGCGCGGATCAC 3052

QY 1921 CTGAGGTGAGGATTCAAACAGCCTGSCCAACATGTGAACCCCTCTCTACTATAAA 1980
      |||...|||
DB 3053 --GAGTTCAGAGATTCGAGACCATCCCGGTAAACCGGTGAACCCCGCTCTACTATAAA 3110

QY 1981 ATACAAAAATTAGTAAGTGTGTTGGCGCATGCTGTAAATCCAGCTACTTGGAGGGT 2040
      |||...|||
DB 3111 ATACAAAAATTAGCGGCGTAGTGGCGGCGCTGTAGTCCAGCTACTTGGAGGCT 3170

QY 2041 GAGGAGGAGAAATTTCTTAAACCCGGGAGGAGAGTTGCAAGTGAAGCGAGATCACACA 2100
      |||...|||
DB 3171 GAGGAGGAGAAATTTCTTAAACCCGGGAGGAGAGTTGCAAGTGAAGCGAGATCACACA 3230

QY 2101 CTGCACTCCAGCTGGGGAGAGAGCGAGACTTCTCTCAAAAAAACAACAAAAAGA 2160
      |||...|||
```

```
DB 3231 CTGCACTCCAGCTGGGCGACAGAGCAGACTCCGCTCTCAAAAAAACAACAAAAA 3290
QY 2161 ATTA 2165
DB 3291 AAAAA 3295

RESULT 12
US-08-724-394A-22
; Sequence 22: Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
;
US-08-724-394A-22

Query Match          7.3%; Score 301.8; DB 2; Length 246240;
Best Local Similarity 68.0%; Pred. No. 2.4e-63;
Matches 493; Conservative 0; Mismatches 177; Indels 55; Gaps 3;

QY 1441 TTTT...TTTGTGAGAAAGACTTGTCTTTTCCACCTGGCTGGAGGCGAGTGGTG 1500
      |||...|||
DB 2626 TTATTTTATTTTGTAGAGCTTGTCTACCCCTGCGCCAGGCTGGAGTGCAGTGGTG 2685

QY 1501 TGAATCTAGCTCAGTCAACTTTCGCCCTCCCGGGTTCGAAGCAATTCCTCGCTTCAGCCT 1560
      |||...|||
DB 2686 CAGTCTTGCTCAGTCAATCCCAAGTTCGAAGCAATTCCTCGCTTCAGCCT 2745

QY 1561 CCCAAGTAGCTGGGACTCAGGCACTTCCACCATGCCAGTGAATTTTGTATTTTATAG 1620
      |||...|||
```

```
Db 2746 CCCAAGCAGCTGGGATTACAGTGCTGCCACCATGCCAGCTAAATTTTGTATTTTAG 2805
QY 1621 TAGAGATGGGATTTCACATTTTGGCCAAAGCTGCTCTCAACATTTTGTGCTCATATTTG 1680
Db 2806 TAGAGACAGGTTTCGCGGTGTTGGCCAGGCTGCTCTCGAACT----- 2848
QY 1681 TTGTAACATTTTCCCTTTTCTGAGTAGGGCCCCCAGACCAAAAAATAAATCTTAG 1740
Db 2849 ----- 2873
QY 1741 AATCCAAATCAGTGTGTTGGTTGACCACTGTCTACTTTGAGAACACAGTGTGACAGGC 1800
Db 2874 TCGGCTCCCAAGATGCTGGGATTACAGGTGTGAGCACTGCCCCAGCGGGTTTCATC 2933
QY 1801 CTCAGGAGTAGGTGATCTCTGCTCGAAGAGAAATAGAAATATATCTCCGGGCC 1860
Db 2934 CTTAAATA-CATACATATAGATATAGATTTCTGTTTATCTAAAGTCTTTATAAGGCC 2992
QY 1861 AGCGTGTGCTCATGCTGTAAATCCAGCACATTTGGGAGGCCAAGGCATGTGATCAC 1920
Db 2993 GGGCGCGTGGCTCACGCCGTGAATCCAGCACATTTGGAGGCCGAGCGGGGATCAC 3052
QY 1921 CTGAGGTCAGAGTTCAAACACCGCTGGCCACATGGTGAACCCCGTCTCTACTAAA 1980
Db 3053 --GAGGTCAGAGTACAGACCATCCCGCTAAACCGTGAACCCCGTCTCTACTAAA 3110
QY 1981 ATCAAAAAATTAGCTAAGTGTGGTGGCGCATGCTCTGAATCCAGCTACTTGGGAGGCT 2040
Db 3111 ATCAAAAAATTAGCGCGGCTAGTGGCGGCGCTCTAGTCCAGCTACTTGGAGGCT 3170
QY 2041 GAGCAGAGAAATTTCTGAACCGGGGAGGCGAGAGTTGTCAGTGAAGCAGATCACACCA 2100
Db 3171 GAGCAGAGAAATGGCGTGAACCGGGGAGGCGAGCTTGCAGTGAGCGGAGATCCCGCCA 3230
QY 2101 CTGCATCCAGCTGGGGGAGAGAGAGAGACTTCCCTCTCAAAAAACAAAAACAAAGA 2160
Db 3231 CTGCATCCAGCTGGGGGAGAGAGAGACTTCCCTCTCAAAAAACAAAAACAAAGA 3290
QY 2161 ATTAA 2165
Db 3291 AAAAA 3295

RESULT 13
US-09-851-896-3
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEPENDENT)
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851,896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

Query Match 7.2%; Score 297.2; DB 4; Length 70000;
Best Local Similarity 68.4%; Pred. No. 1.8e-62;
Matches 491; Conservative 0; Mismatches 208; Indels 19; Gaps 5;
```

```
QY 1441 TTTTATTTTATTTTGTGAGAAAGAGTCTGCTCTT-TCACCTCGGCTGGAGGCGAGTGGT 1499
Db 49703 TTTATTTATTTTGTGAGACAGAGTTTCGCTTTTGTGCGGAGCTGGAGTGCATGGC 49762
QY 1500 GTGATCTCAGCTCACTGCAACTTTGGCTCCCGGTTTCAAGCAATTTCTCTGCTTCAGCC 1559
```

```
Db 49763 ACGAICTGGCTCACCGAACCTCTGCTCTCTGGGTTCAAGCGATTCCTCGCTCAGCT 49822
QY 1560 TCCCAAGTAGTGGGACTACAGGCACATTCACCATGCCCAGCTGATTTTGTATTTTAA 1619
Db 49823 TCCGAGTAGTGGGANTTACGTTTCATGCGCCACACACCTGGCT-AAATTTGTATTTTA 49881
QY 1620 GTAGAGATGGGATTTCACTTTGTTGGCCAGCTGCTCTCAAACTTTTGTGCTCATATTT 1679
Db 49882 GTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTTGAACCT-----CCTGACCT 49933
QY 1680 GTTGTAACTATTGTCCTTTGCTGAGGTAGGGCCCCCAGACCAAAAAATAAATCTTA 1739
Db 49934 CAGGTGTTCCACCTGCTCAGCTCCCAAACTGCTGGGATTACAGGCTGAGCCACCGTG 49993
QY 1740 GAATCCAAATCAGTGTGTTGGTTTGACCACTGTCTACTTTGAGAACACACAGTGTGACAGGG 1799
Db 49994 CCGGCTGGCTTAACTTAACTGACTTAAATTTATTTTGAAGAAATGTT-----G 50046
QY 1800 CCTCAGGAGTAGAGTGTCTCTGCTCGAAGAGAAATAGAAATATTTCTCCGGGC 1859
Db 50047 CATCACTACCATAAATGGCCATCCATCCAGTGCACCAACCATATAAAGGTAAACCAATGC 50106
QY 1860 CAGCGCTGGTGCCTGTAAATCCAGCACTTTGGGAGGCCAAGGCATGTGGATCA 1919
Db 50107 AGGCGCGGTGGCTCACGCGCTGAATCCAGCATTTTGGGAGGCCGAGTGGCGAATCA 50166
QY 1920 CTGAGTCAAGGAGTTCAAAACACAGCTGGCCAACTGGTGAACCCCGTCTCTACTAAA 1979
Db 50167 C--GAGGTCAGGAGTTCAAGACACCGCTGGCCACATGGTGAACCCCGTCTCTACTAAA 50224
QY 1980 AATACAAAAATTAGCTAAGTGTGGTGGCGCATGCCCTGTAATCCAGCTACTTGGGAGGG 2039
Db 50225 AATACAAAAATTAGCTGGCGGTAGTGGTGGCCACCTGTAATFACCAGCTACTCTGGAGGC 50284
QY 2040 TGAGGCGAGGAAATTTCTTGAACCCCGGAGGCGAGAGTTGTCAGTGAAGCGAGATCACACC 2099
Db 50285 TGAGGCGAGGAAATTTCTTGAACCTCGGTAGGTAGAGTTGCAGTGGCTGAGATCAGGCC 50344
QY 2100 ACTGCACTCCAGCTGGGGGAGAGAGAGACTTCTCTCAAAAAACAAAAACAAA 2157
Db 50345 ACTGCACTCCAGCTGGGTGACAGAGTGGCTCCCGGTTCTCAAAAAACAAAAACAAA 50402

RESULT 14
US-09-813-817-3
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001178
; CURRENT APPLICATION NUMBER: US/09/813,817
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
US-09-813-817-3

Query Match 7.1%; Score 296.6; DB 4; Length 59065;
Best Local Similarity 68.2%; Pred. No. 2.3e-62;
Matches 518; Conservative 0; Mismatches 224; Indels 17; Gaps 7;
```

```
QY 1441 TTTTATTTTATTTTGTGAGAAAGAGTCTGCTCT-TCACCTCGGCTGGAGGCGAGTGGT 1499
Db 11642 TTGTTTGTGTTTGTGAGACAGTCTCACTCTGTCTCCCAAGGCTGGAGTGCATGGC 11701
QY 1500 GTGATCTCAGCTCACTGCAACTTTTCGCCCTCCCGGTTTCAAGCAATTTCTCTGCTTCAGCC 1559
```

```
Db 11702 ACAATTCGGCTCACTGCAACTTCTGCTCCCGGGTTTAAGCAATTCCTCTGCTCAGCC 11761
Qy 1560 TCCCAAGTAGCTGGGACTAGAGCACTTCCACCAATGCCCGGAGATTTTGTATTTTA 1619
Db 11762 TCCCAAGTAGCTGGGATTA-AGGCACCGCGCCCATACACCACTAAATTTTATATTTTA 11820
Qy 1620 GTAGAGATGGGATTTCACTTTTGGCCCAAGCTGTGTCFCAAACTTTTGGCTGTCATATT 1679
Db 11821 GTAGAGATGGGTTTGGCCATGTTGGCCCAAGCTGTGTCFCAAACTTTTGGCTGTCATATT 11880
Qy 1680 ---GTGTAACATAATGTTTCTTTTCTGAGTAGGCGCCCGCCAGACCAAAAAATAAATC 1736
Db 11881 CCACCTGCTCGGCTCCCAAAGTGTGGGATTTACAGGTGTGAGCCACTGCACCTGGCTG 11940
Qy 1737 TTAGATCCCAATCAGTGTGTGTTGACCACTGTCTACTTTGAGAACCACTGAGTGA 1796
Db 11941 GAAGGAGTGATCTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 12000
Qy 1797 GGGCTCAGGAGTAGAGGTGATC-----TCTGCTCGAAAGAGAAATAGAAATGAAAAAT 1848
Db 12001 GTCCTCTCTACCTTGTATCTTCCAACTTCCAGTGTCTTGTGATTAAGACCAAAATC 12060
Qy 1849 ATTCTCCGGGCGGCGGTGGTCTGATGCTGTAATCCAGCACTTTGGGAGGCGCAAGG 1908
Db 12061 CTTAACTTGGCGAGCGCGGTGGCTCAACCTATCATCTCAGCACTTTGGGAGGCGGAGG 12120
Qy 1909 CATGTGATCACCTGAGTCAAGGAGTTCAAACAGAGCTGCGCAACATGTTGAACCCCA 1968
Db 12121 CAGGAGATCA--TGAAGTCAAGAGATTGAGACCATCTCGCCCAACATGGTGAACCCCA 12178
Qy 1969 TCTCTACTAAAAATACAAAAAATAGCTTAAGTGTGGTGGCGCATGCCCTGTATCCAGCT 2028
Db 12179 TCTCTACTAAAAATAC--AAAAATAGCTGTGGTGGCGCTGTGCTGTAGTCCAGCT 12237
Qy 2029 ACTTGGAGGTGAGGAGAGAAATTTCTTGAACCCGGGAGGAGAGGTTGCAAGTGAAGC 2088
Db 12238 ACTTGGAGGTGAGGAGGAGAAATCACTTGAACCTGGGAGGAGAGGTTGCAAGTGAAGC 12297
Qy 2089 GAGATCACCACTGCACTCCAGCTGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2148
Db 12298 GAGATCACCACTGCACTCCAGCT--GGTACAGAGTAAGAGTCCATCTCAAAAAAATA 12356
Qy 2149 AAAAAACAAAGAAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTA 2187
Db 12357 AAAAAACAAAGAAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTA 12395
```

RESULT 15

```
US-978-197-3
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001178DIV
; CURRENT APPLICATION NUMBER: US/09/978,197
; CURRENT FILING DATE: 2001-10-17
; PRIOR APPLICATION NUMBER: 09/813,817
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-978-197-3
```

```
Query Match 97.1%; Score 296.6; DB 4; Length 59065;
Best Local Similarity 68.2%; Pred. No. 2.3e-62;
Matches 518; Conservative 0; Mismatches 224; Indels 17; Gaps 7;
```

```
Qy 1441 TTTTCTTTTCTTTTGTGAGAAAGAGTCTTGCTCT--TTCACCTCTGGCTGGAGGCAAGTGT 1499
Db 11642 TTGTTTTTCTTTTGTGAGAAAGAGTCTTGCTCT--TTCACCTCTGGCTGGAGGCAAGTGT 11701
Qy 1500 GTGATCTCAGCTCACTGCAACTTTTCCCTCCCGGGTTTCAAGCAATTTCTCTGCTTTCAGCC 1559
Db 11702 ACAATTCGGCTCACTGCAACTTCTGCTCCCGGGTTTAAAGCAATTTCTCTGCTTTCAGCC 11761
Qy 1560 TCCCAAGTAGCTGGGACTAGAGCACTTCCCAACATGCCCGGAGATTTTGTATTTTA 1619
Db 11762 TCCCAAGTAGCTGGGATTA-AGGCACCGCGCCCATACACCACTAAATTTTATATTTTA 11820
Qy 1620 GTAGAGATGGGATTTCACTTTTGGCCCAAGCTGTGTCFCAAACTTTTGGCTGTCATATT 1679
Db 11821 GTAGAGATGGGTTTGGCCATGTTGGCCCAAGCTGTGTCFCAAACTTTTGGCTGTCATATT 11880
Qy 1680 ---GTGTAACATAATGTTTCTTTTCTGAGTAGGCGCCCGCCAGACCAAAAAATAAATC 1736
Db 11881 CCACCTGCTCGGCTCCCAAAGTGTGGGATTTACAGGTGTGAGCCACTGCACCTGGCTG 11940
Qy 1737 TTAGATCCCAATCAGTGTGTGTTGACCACTGTCTACTTTGAGAACCACTGAGTGA 1796
Db 11941 GAAGGAGTGATCTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 12000
Qy 1797 GGGCTCAGGAGTAGAGGTGATC-----TCTGCTCGAAAGAGAAATAGAAATGAAAAAT 1848
Db 12001 GTCCTCTCTACCTTGTATCTTCCAACTTCCAGTGTCTTGTGATTAAGACCAAAATC 12060
Qy 1849 ATTCTCCGGGCGGCGGTGGTCTGATGCTGTAATCCAGCACTTTGGGAGGCGCAAGG 1908
Db 12061 CTTAACTTGGCGAGCGCGGTGGCTCAACCTATCATCTCAGCACTTTGGGAGGCGGAGG 12120
Qy 1909 CATGTGATCACCTGAGTCAAGGAGTTCAAACAGAGCTGCGCAACATGTTGAACCCCA 1968
Db 12121 CAGGAGATCA--TGAAGTCAAGAGATTGAGACCATCTCGCCCAACATGGTGAACCCCA 12178
Qy 1969 TCTCTACTAAAAATACAAAAAATAGCTTAAGTGTGGTGGCGCATGCCCTGTATCCAGCT 2028
Db 12179 TCTCTACTAAAAATAC--AAAAATAGCTGTGGTGGCGCTGTGCTGTAGTCCAGCT 12237
Qy 2029 ACTTGGAGGTGAGGAGGAGAAATTTCTTGAACCCGGGAGGAGAGGAGGAGGAGGAGGAG 2088
Db 12238 ACTTGGAGGTGAGGAGGAGAAATCACTTGAACCTGGGAGGAGAGGAGGAGGAGGAGGAG 12297
Qy 2089 GAGATCACCACTGCACTCCAGCTGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2148
Db 12298 GAGATCACCACTGCACTCCAGCT--GGTACAGAGTAAGAGTCCATCTCAAAAAAATA 12356
Qy 2149 AAAAAACAAAGAAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTA 2187
Db 12357 AAAAAACAAAGAAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTA 12395
```

Search completed: May 20, 2003, 19:56:27
Job time : 3074 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 19:06:28 ; Search time 489 Seconds
(without alignments)
11206.377 Million cell updates/sec

Title: US-10-016-725-15

Perfect score: 4150
Sequence: 1 aggaattcatcatttaaat.....tgagttccagcctggccc 4150

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4150	100.0	4150	12	US-10-016-725-15
C 2	429	10.3	17704	9	US-10-092-154-1865
C 3	429	10.3	17704	10	US-09-764-847-1865
C 4	427.4	10.3	17707	9	US-10-092-154-1864
C 5	427.4	10.3	17707	10	US-09-764-847-1864
C 6	409.8	9.9	5648	9	US-09-981-353-183
C 7	409.8	9.9	5648	10	US-09-919-172-70
C 8	349	8.4	13216	9	US-09-764-891-7810
C 9	347.4	8.4	29163	9	US-09-764-891-7809
C 10	342.8	8.3	26928	9	US-10-020-141-7
C 11	342.8	8.3	26928	10	US-09-880-107-2278
C 12	329.6	7.9	9372	10	US-09-764-877-3049
C 13	329.6	7.9	9372	9	US-09-764-877-3244
C 14	327.6	7.9	17397	9	US-10-091-504-1945
C 15	327.6	7.9	17397	10	US-09-764-869-1945
C 16	327.6	7.9	19334	9	US-10-091-504-1943
C 17	327.6	7.9	19334	10	US-09-764-869-1943
C 18	327.6	7.9	19345	9	US-10-091-504-1944
C 19	327.6	7.9	19345	10	US-09-764-869-1944

```

c 20 327.6 7.9 25423 9 US-10-072-349-193 Sequence 193, App
c 21 327.6 7.9 25423 10 US-09-764-855-193 Sequence 193, App
c 22 327 7.9 5932 9 US-09-764-891-6396 Sequence 6396, Ap
c 23 326.4 7.9 32148 9 US-09-764-891-6906 Sequence 6906, Ap
c 24 326 7.9 25424 9 US-10-072-349-194 Sequence 194, App
c 25 326 7.9 25424 10 US-09-764-855-194 Sequence 194, App
c 26 325.4 7.8 32145 9 US-09-764-891-7319 Sequence 7319, Ap
c 27 324.4 7.8 32191 9 US-09-764-891-6304 Sequence 6304, Ap
c 28 324 7.8 3471 9 US-10-092-154-1893 Sequence 1893, Ap
c 29 324 7.8 3471 10 US-09-764-847-1893 Sequence 1893, Ap
c 30 323.2 7.8 32134 9 US-09-764-891-6303 Sequence 6303, Ap
c 31 321.8 7.8 32216 9 US-09-764-891-9613 Sequence 9613, Ap
c 32 320.4 7.7 465237 10 US-09-933-267A-1 Sequence 1, Appli
c 33 320.2 7.7 7703 9 US-10-073-961-415 Sequence 415, App
c 34 320.2 7.7 7703 9 US-10-073-961-415 Sequence 415, App
c 35 320.2 7.7 7703 10 US-09-764-887-415 Sequence 415, App
c 36 320.2 7.7 7703 10 US-09-764-887-456 Sequence 456, App
c 37 319.6 7.7 32192 9 US-10-092-154-1416 Sequence 1416, Ap
c 38 319.6 7.7 32192 9 US-09-764-891-7945 Sequence 7945, Ap
c 39 319.6 7.7 32192 9 US-09-764-891-8220 Sequence 8220, Ap
c 40 319.6 7.7 32192 9 US-09-764-891-8319 Sequence 8319, Ap
c 41 319.6 7.7 32192 10 US-09-764-847-1416 Sequence 1416, Ap
c 42 319.6 7.7 65608 9 US-09-954-531-180 Sequence 180, App
c 43 319.6 7.7 65608 10 US-09-962-436-292 Sequence 292, App
c 44 319.6 7.7 65608 10 US-09-962-832-119 Sequence 119, App
c 45 317.8 7.7 10527 9 US-10-116-420-100 Sequence 100, App

```

ALIGNMENTS

RESULT 1

```

US-10-016-725-15
; Sequence 15, Application US/10016725
; Patent No. US20020151018A1
; GENERAL INFORMATION:
; APPLICANT: Prouty, Stephen
; APPLICANT: Zhang, Lin
; TITLE OF INVENTION: Stearyl-CoA Desaturase Promoter
; FILE REFERENCE: J6J2065
; CURRENT APPLICATION NUMBER: US/10/016,725
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 4150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-725-15

```

Query Match 100.0%; Score 4150; DB 12; Length 4150;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGGAATTCATCCATTTAAATCATACAAATTAATGGCTTTTAGTATATATCATAGTTGTGC 60
|
Db 1 AGGAATTCATCCATTTAAATCATACAAATTAATGGCTTTTAGTATATATCATAGTTGTGC 60
|
QY 61 ATCCATCACAAATCCATTTTAAACAGTTTATCTCCAAATTAATAACCTGGATTCCTT 120
|
Db 61 ATCCATCACAAATCCATTTTAAACAGTTTATCTCCAAATTAATAACCTGGATTCCTT 120
|
QY 121 AGGCATCACCCCCCAACATCCTCCATCCTCTCCAGCCCTGGCAACCAACCAATCTAC 180
|
Db 121 AGGCATCACCCCCCAACATCCTCCATCCTCTCCAGCCCTGGCAACCAACCAATCTAC 180
|
QY 181 TTCTGTCTCTATAAAATTTGCCAATTTCTGGACATTTCTGACATTTATATAATGAAGCAACAACAT 240
|
Db 181 TTCTGTCTCTATAAAATTTGCCAATTTCTGGACATTTCTGACATTTATATAATGAAGCAACAACAT 240
|
QY 241 GTGAGACTTTGTGACAGCTGGCTTTCACITTAGCATTCATTTTAAAGCTCATTAATCTTA 300
|

```

Db 241 GTGAGACTTTGTGACTGGCTGCTTTCACCTTAGCATTTCTATTTTAAAGGCTCATTTATGTTA 300
QY 301 CAGTACTTAGCAGTACTTTCATTTCTTTTATTCATAAGTATTCACGTGTGGGGTAT 360
Db 301 CAGTACTTAGCAGTACTTTCATTTCTTTTATTCATAAGTATTCACGTGTGGGGTAT 360
QY 361 CCCATATCATATTATTAGACAGAGTCTCCTACTGTGCACCCAGGCTGGAGTGCAGTGGC 420
Db 361 CCCATATCATATTATTAGACAGAGTCTCCTACTGTGCACCCAGGCTGGAGTGCAGTGGC 420
QY 421 ACATATAGCTACTGTAACTCAAACTCCTGGGCTCAAGTGATTCCTACTACCTCAGCC 480
Db 421 ACAAATAGCTACTGTAACTCAAACTCCTGGGCTCAAGTGATTCCTACTACCTCAGCC 480
QY 481 TCCAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGGCTAAATTTTTTTTTTAAT 540
Db 481 TCCAGAGTAGCTAGGACTACAGGCACACAGCCATACCTGGGCTAAATTTTTTTTTTAAT 540
QY 541 TTTCAATTTAAGTATTCATTTTCTTTTCTTTTGTGTTGTGTTTGTAGATAGGGTCTC 600
Db 541 TTTCAATTTAAGTATTCATTTTCTTTTCTTTTGTGTTGTGTTTGTAGATAGGGTCTC 600
QY 601 ACTTTGTACCCAGCTGGAGGAGTGGCATGGTGCACAGCTGAGAGCCTTGACTTTCCT 660
Db 601 ACTTTGTACCCAGCTGGAGGAGTGGCATGGTGCACAGCTGAGAGCCTTGACTTTCCT 660
QY 661 GGGCTCAAGTGATCCTCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGCTGTCA 720
Db 661 GGGCTCAAGTGATCCTCCTGCTCAGCCTCCCAAGTAGCTGGGACTACAAACAGCTGTCA 720
QY 721 CCATGCTGGCTGATATTTTCTTTTCTTTTGTGTTGTGTTTGTAGATAGGGTGGG 780
Db 721 CCATGCTGGCTGATATTTTCTTTTCTTTTGTGTTGTGTTTGTAGATAGGGTGGG 780
QY 781 GTACAGTGGGCTAATAATAGCTACAGCTCCGCTCCCTCCTGGCTCAAGCAATCCGCTG 840
Db 781 GTACAGTGGGCTAATAATAGCTACAGCTCCGCTCCCTCCTGGCTCAAGCAATCCGCTG 840
QY 841 GCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTTGTGCCACAGGCCAGCTAAAGTTT 900
Db 841 GCCTCAGCATCCTGAGTAGCTGGGACTACAGGCTTGTGCCACAGGCCAGCTAAAGTTT 900
QY 901 AAAAAATGATTTTGTATAGAGAGTCTTGTATGTTGCTCAGGCTGATTTTATG 960
Db 901 AAAAAATGATTTTGTATAGAGAGTCTTGTATGTTGCTCAGGCTGATTTTATG 960
QY 961 TTGAGACAAGGTCTCAGTATGTTGCCATGATCCCCCAGCTCCACTTCCCAAAGTGCTCA 1020
Db 961 TTGAGACAAGGTCTCAGTATGTTGCCAIGATPCCCCCAGCTCCACTTCCCAAAGTGCTCA 1020
QY 1021 TCTTATCTGTTTCATTAGTCAAGTTGACAGACATTTAGTGTGTTTCCACTTTTTCACCATTA 1080
Db 1021 TCTTATCTGTTTCATTAGTCAAGTTGACAGACATTTAGTGTGTTTCCACTTTTTCACCATTA 1080
QY 1081 TGAATATCTCAGTGAATATTCATGATACATTTGTGTTGGCATATGTTTTCATTTCT 1140
Db 1081 TGAATATCTCAGTGAATATTCATGATACATTTGTGTTGGCATATGTTTTCATTTCT 1140
QY 1141 GTTGGGTTTATATCTAGGAGTGAATTTGCTGGATCCCGGGTAAATTTTTCACAGGCAGAG 1200
Db 1141 GTTGGGTTTATATCTAGGAGTGAATTTGCTGGATCCCGGGTAAATTTTTCACAGGCAGAG 1200
QY 1201 TTCAGGGGAAGAAAACTTGGGAAAAATGAAGCATGTTTGAATAATCAGCAAGTGCAGGG 1260
Db 1201 TTCAGGGGAAGAAAACTTGGGAAAAATGAAGCATGTTTGAATAATCAGCAAGTGCAGGG 1260
QY 1261 GTTTTTCGGAGTTTATTTATATTCGTTGACAAATGTCGATTTGATGAAGATACAAG 1320
Db 1261 GTTTTTCGGAGTTTATTTATATTTCTGTTGACAAATGTCGAGTTTGTATGAAGATACAAG 1320
QY 1321 TTATACTAAGTGAAGTGAAGTTAAGGCTGGAATAGGCGCTTCAGAGTAAAATCATGA 1380
Db 1321 TTATACTAAGTGAAGTGAAGTTAAGGCTGGAATAGGCGCTTCAGAGTAAAATCATGA 1380

QY 1381 AGCACTTTGAATACCAAAATTAAGAGCTTGGCTGTAAACAAAAATAATAAAAAATCACAA 1440
Db 1381 AGCACTTTGAATACCAAAATTAAGAGCTTGGCTGTAAACAAAAATAATAAAAAATCACAA 1440
QY 1441 TTTTTTTTTTTTTTTTGAGAAAGAGTCTGCTCTTTTCCCTTCCCTGGCTGGAGGCGAGTGGTG 1500
Db 1441 TTTTTTTTTTTTTTTTGAGAAAGAGTCTGCTCTTTTCCCTTCCCTGGCTGGAGGCGAGTGGTG 1500
QY 1501 TGAFTCTCAGCTCACTGCAAACTTTTCGCTCCCGGGTTTCAAGCAATTCCTCTGCTTCAGCCT 1560
Db 1501 TGAFTCTCAGCTCACTGCAAACTTTTCGCTCCCGGGTTTCAAGCAATTCCTCTGCTTCAGCCT 1560
QY 1561 CCCAAGTAGCTGGGACTACAGGCACITTCACCAATGCCCAGCTGATTTTGTATTTTAG 1620
Db 1561 CCCAAGTAGCTGGGACTACAGGCACITTCACCAATGCCCAGCTGATTTTGTATTTTAG 1620
QY 1621 TAGAGATGGGATTTTCACTTTTGTGGCAAGCTGTCTCAAACTTTTGTCTGCATAATTG 1680
Db 1621 TAGAGATGGGATTTTCACTTTTGTGGCAAGCTGTCTCAAACTTTTGTCTGCATAATTG 1680
QY 1681 TTGTAACATATTGTTCCCTTTTGTGCTGAGGTAGGGCCCCCAGACCAAAAAATAAATCCTTAG 1740
Db 1681 TTGTAACATATTGTTCCCTTTTGTGCTGAGGTAGGGCCCCCAGACCAAAAAATAAATCCTTAG 1740
QY 1741 AATCCAAATCAGTGTGTTGTTGACCACTGTCACTTTGAGAACCCACAGTGTGACCAGGGC 1800
Db 1741 AATCCAAATCAGTGTGTTGTTGACCACTGTCACTTTGAGAACCCACAGTGTGACCAGGGC 1800
QY 1801 CTCAGGAGTAGAGTGTATCTGCTCGAAAAGAGAAATAGAAATGAAATATTTCTCCGGGCC 1860
Db 1801 CTCAGGAGTAGAGTGTATCTGCTCGAAAAGAGAAATAGAAATGAAATATTTCTCCGGGCC 1860
QY 1861 AGCGTGGTGGCTCATGCCCTGTATTCACAGACTTTTGGAGGCCAAGCATGTGGATCAC 1920
Db 1861 AGCGTGGTGGCTCATGCCCTGTATTCACAGACTTTTGGAGGCCAAGCATGTGGATCAC 1920
QY 1921 CTGAGGTTCAGAGTTCAAAACCCAGCCTGGCCAAACATGTTGAAACCCCTCTACTAAAA 1980
Db 1921 CTGAGGTTCAGAGTTCAAAACCCAGCCTGGCCAAACATGTTGAAACCCCTCTACTAAAA 1980
QY 1981 ATACAAAAATTAGTCAAGTGTGGTGGCGATGCTGTAAATCCAGCTACTTGGGAGGGT 2040
Db 1981 ATACAAAAATTAGTCAAGTGTGGTGGCGATGCTGTAAATCCAGCTACTTGGGAGGGT 2040
QY 2041 GAGCAGAGCAATTTCTTTGAACCCGGGAGGCGAGGTTGCAGTGAAGCGAGATCACACCA 2100
Db 2041 GAGCAGAGCAATTTCTTTGAACCCGGGAGGCGAGGTTGCAGTGAAGCGAGATCACACCA 2100
QY 2101 CTGCACCTCCAGCCTGGGGGAGAGCGAGACTTCTCTCAAAAAACAAAAACAAAAAGA 2160
Db 2101 CTGCACCTCCAGCCTGGGGGAGAGCGAGACTTCTCTCAAAAAACAAAAACAAAAAGA 2160
QY 2161 ATTAAGCAAAATTAGACATTCGAGAGAGAACCTGAAGGGGTTCAGACCATCAGATTTTC 2220
Db 2161 ATTAAGCAAAATTAGACATTCGAGAGAGAACCTGAAGGGGTTCAGACCATCAGATTTTC 2220
QY 2221 TGTGCCACATGCCCAAGTACTTCTGAGGCATGACTTGGATGAGCTGTCCACATCTGAATCA 2280
Db 2221 TGTGCCACATGCCCAAGTACTTCTGAGGCATGACTTGGATGAGCTGTCCACATCTGAATCA 2280
QY 2281 TCCAGTCTTCTTCAGAACTTTTCACCGGACAGGAGCCAGGACTGGAATGCAAGTCTCCT 2340
Db 2281 TCCAGTCTTCTTCAGAACTTTTCACCGGACAGGAGCCAGGACTGGAATGCAAGTCTCCT 2340
QY 2341 GGTCACTGCCAGAGAGTGTGGCCTTGACCCCTGAGACCAAGTGGCCAAACAAAGGAGCTGCT 2400
Db 2341 GGTCACTGCCAGAGAGTGTGGCCTTGACCCCTGAGACCAAGTGGCCAAACAAAGGAGCTGCT 2400
QY 2401 AGTCTACCTCCAGAGAAATCCCAAGTGTCTTCTTCCCTGGGAAAGTGAATCATTTGGGCGAG 2460
Db 2401 AGTCTACCTCCAGAGAAATCCCAAGTGTCTTCTTCCCTGGGAAAGTGAATCATTTGGGCGAG 2460

```
QY 2461 CACTCCGTAATTTCTCCTCTTCCAGGGAAGGATCCTAGGCGAGTATTTGGGAAAGACA 2520
|
|
|
Db 2461 CACTCCGTAATTTCTCCTCTTCCAGGGAAGGATCCTAGGCGAGTATTTGGGAAAGACA 2520
|
|
|
QY 2521 TGGGCATGGAAGACACCGGGTGAAATGCAATAGCGTCCCTGGTCTGAGCTCTCATGTTAA 2580
|
|
|
Db 2521 TGGGCATGGAAGACACCGGGTGAAATGCAATAGCGTCCCTGGTCTGAGCTCTCATGTTAA 2580
|
|
|
QY 2581 GGCTCCACAGACACGAAAGATGSGGCGACAGGACAGATAGTGGTTCAGAGCATC 2640
|
|
|
Db 2581 GGCTCCACAGACACGAAAGATGSGGCGACAGGACAGATAGTGGTTCAGAGCATC 2640
|
|
|
QY 2641 TCAGGACCGAGGCAATATGTTCTCTGAGCAGGATTAAGAGCTTTGGCTCTCATATGGT 2700
|
|
|
Db 2641 TCAGGACCGAGGCAATATGTTCTCTGAGCAGGATTAAGAGCTTTGGCTCTCATATGGT 2700
|
|
|
QY 2701 GTTTCTGGCTCACTGCCAGCTCCGTCACTTACTGTTGCTGAGCATCGGGCAAGTTA 2760
|
|
|
Db 2701 GTTTCTGGCTCACTGCCAGCTCCGTCACTTACTGTTGCTGAGCATCGGGCAAGTTA 2760
|
|
|
QY 2761 TTCCATCTCTCCATATCTCTTCTCTCACTTTTAAATGGGAATATGGGTACCCACCTCC 2820
|
|
|
Db 2761 TTCCATCTCTCCATATCTCTTCTCTCACTTTTAAATGGGAATATGGGTACCCACCTCC 2820
|
|
|
QY 2821 CAGGGTCACAGAGAGGCTTACAGAAACGATTTCTGTGAATTTGGCTTGCAGTAATAATC 2880
|
|
|
Db 2821 CAGGGTCACAGAGAGGCTTACAGAAACGATTTCTGTGAATTTGGCTTGCAGTAATAATC 2880
|
|
|
QY 2881 AATACCTGCCAGCTATTTCTTATTCACATCAAGCCCTTTGCGCTGCTGCTGGTGAAAA 2940
|
|
|
Db 2881 AATACCTGCCAGCTATTTCTTATTCACATCAAGCCCTTTGCGCTGCTGCTGGTGAAAA 2940
|
|
|
QY 2941 CACATGTCAGTGTCTTCTGAGGCTTTCCACAAAGAAATTCACAAATTAACCTGCCAG 3000
|
|
|
Db 2941 CACATGTCAGTGTCTTCTGAGGCTTTCCACAAAGAAATTCACAAATTAACCTGCCAG 3000
|
|
|
QY 3001 TCTGAGAATCTCCAAACATCTCCGACAGCATCTGAGGCGGGCTTGGGATGGGAC 3060
|
|
|
Db 3001 TCTGAGAATCTCCAAACATCTCCGACAGCATCTGAGGCGGGCTTGGGATGGGAC 3060
|
|
|
QY 3061 TGCCCGCCGGGTCCTGAACAGATGGTGGCGCAGGACACACACACACAGCCAGCTG 3120
|
|
|
Db 3061 TGCCCGCCGGGTCCTGAACAGATGGTGGCGCAGGACACACACACAGCCAGCTG 3120
|
|
|
QY 3121 TGTGTGGCGCGGAGTCCGCTGCGGTCCCGGTGAGCAGCGCTGCTGGTGGGCGGGGC 3180
|
|
|
Db 3121 TGTGTGGCGCGGAGTCCGCTGCGGTGAGCAGCGCTGCTGGTGGGCGGGGC 3180
|
|
|
QY 3181 AGAGCCATTGTTGCGAGCGGTACCGAGCCCGCCGCGCTGCGCGGAGGAGGCGGGCT 3240
|
|
|
Db 3181 AGAGCCATTGTTGCGAGCGGTACCGAGCCCGCCGCGCTGCGCGGAGGAGGCGGGCT 3240
|
|
|
QY 3241 TCCGCGCTCCCAAGCTCCAGATCTCTGGGTGGCTGCGCAGCTCTCCCTGCCACGCGCTG 3300
|
|
|
Db 3241 TCCGCGCTCCCAAGCTCCAGATCTCTGGGTGGCTGCGCAGCTCTCCCTGCCACGCGCTG 3300
|
|
|
QY 3301 GGGGAGCGGAAGACGCGGAGATTTAGTGGTGGGCGCCCGCCCGAGGGTTACCACT 3360
|
|
|
Db 3301 GGGGAGCGGAAGACGCGGAGATTTAGTGGTGGGCGCCCGCCCGAGGGTTACCACT 3360
|
|
|
QY 3361 GTTTCTTGAGAAATCTCCCAAGTGGCCACCCAGCCGCTTCTCGTGTGCCGAGGCGCGGT 3420
|
|
|
Db 3361 GTTTCTTGAGAAATCTCCCAAGTGGCCACCCAGCCGCTTCTCGTGTGCCGAGGCGCGGT 3420
|
|
|
QY 3421 CCTGGGTAGGCTCCGCGCCCGCCAGCCCGGCTCCCGAGCGCTTCCAGAGAGAAA 3480
|
|
|
Db 3421 CCTGGGTAGGCTCCGCGCCCGCCAGCCCGGCTCCCGAGCGCTTCCAGAGAGAAA 3480
|
|
|
QY 3481 GCTCCGACCGGGATCCCGGGCAGAGGCGCCAGCGGGTGGAGAGAAAGCTGAGAAGG 3540
|
|
|
Db 3481 GCTCCGACCGGGATCCCGGGCAGAGGCGCCAGCGGGTGGAGAGAAAGCTGAGAAGG 3540
|
|
|
QY 3541 AGAAACAGAGGGGAGGAGGAGCTGGCGGACAGGGAACACAGATTTCGCCCGCA 3600
|
|
|
```

```
Db 3541 AGAAACAGAGGGGAGGAGGAGCTGGCGCAGAGGAACACAGATTTCGCCCGCA 3600
|
|
|
QY 3601 GCCAATGGCAACGCGCAGGAGGAGTGCGACCAAAATTCCTTTTCGGCCAAATGACGAGCCGGA 3660
|
|
|
Db 3601 GCCAATGGCAACGCGCAGGAGGAGTGCGACCAAAATTCCTTTTCGGCCAAATGACGAGCCGGA 3660
|
|
|
QY 3661 GTTTACAGAGCCCTCATAGCATTTCCCAAGAGCGAGGCGAGAGCGCGGGTGG 3720
|
|
|
Db 3661 GTTTACAGAGCCCTCATAGCATTTCCCAAGAGCGAGGCGAGAGCGCGGGTGG 3720
|
|
|
QY 3721 TGTGTGTCTGCGGTGTCGCGAGCATCCCGGGCGCTGCTGCGGTGCGCGAGCCTCGGCC 3780
|
|
|
Db 3721 TGTGTGTCTGCGGTGTCGCGAGCATCCCGGGCGCTGCTGCGGTGCGCGAGCCTCGGCC 3780
|
|
|
QY 3781 TGTGTCTCTCCCTCCCGCTTACCTCCACGCGGAGCGCGCGCGCCCACTCC 3840
|
|
|
Db 3781 TGTGTCTCTCCCTCCCGCTTACCTCCACGCGGAGCGCGCGCGCCCACTCC 3840
|
|
|
QY 3841 TCGCAGCTTTGGCCCTGCTTGGCAGCGGATAAAGGGGCTGAGGAAATACCGGACAGCT 3900
|
|
|
Db 3841 TCGCAGCTTTGGCCCTGCTTGGCAGCGGATAAAGGGGCTGAGGAAATACCGGACAGCT 3900
|
|
|
QY 3901 CACCCGTTGCCAGCTTAGCCTTTAAATTCGCGCTCGGGACCTTCCACGACCGCGCT 3960
|
|
|
Db 3901 CACCCGTTGCCAGCTTAGCCTTTAAATTCGCGCTCGGGACCTTCCACGACCGCGCT 3960
|
|
|
QY 3961 AGCGCCACAACAGCTAGCTGCAAGCGCGCGGGCTCAGCGGTACCGCGGGCTTCG 4020
|
|
|
Db 3961 AGCGCCACAACAGCTAGCTGCAAGCGCGCGGGCTCAGCGGTACCGCGGGCTTCG 4020
|
|
|
QY 4021 AAACCGCAGTCTCCCGGACCGGAACTCCGCTCCGGAGCTCAGCGCCCTGGAAGTG 4080
|
|
|
Db 4021 AAACCGCAGTCTCCCGGACCGGAACTCCGCTCCGGAGCTCAGCGCCCTGGAAGTG 4080
|
|
|
QY 4081 ATCCCGCATCCGAGAGCAGATGCGGCGCCACTTGTGTCAGGACGATGTGATTTCCC 4140
|
|
|
Db 4081 ATCCCGCATCCGAGAGCAGATGCGGCGCCACTTGTGTCAGGACGATGTGATTTCCC 4140
|
|
|
QY 4141 AGCCTGGGCC 4150
|
|
|
Db 4141 AGCCTGGGCC 4150
|
|
|
RESULT 2
US-10-092-154-1865/c
; Sequence 1865, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092.154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1865
; LENGTH: 17704
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1865
Query Match 10.3%; Score 429; DB 9; Length 17704;
Best Local Similarity 100.0%; Pred. No. 4.4e-104;
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3722 GTGTGTGCGGTGTCGCGAGCATCCCGCGCGCTTGTGTCGCTGCGCGAGCCTTCGSCCT 3781
|
|
|
Db 17704 GTGTGTGCGGTGTCGCGAGCATCCCGCGCGCTTGTGTCGCTGCGCGAGCCTTCGSCCT 17645
|
|
|
QY 3782 CTGTCTCTCTCCCTCCCGCTTACCTCCACGCGGACCGCCCGCGCAGTCACTCT 3841
|
|
|
```


Db 17644 CTGTCCTCCCTCCCGCCCTTACCTCCACGCGGACCGCCCGCGCAGTCAACTCCT 17585
QY 3842 CGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 3901
Db 17584 CGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 17525
QY 3902 ACCGTTTGGCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTCCACGACCGGGCTA 3961
Db 17524 ACCGTTTGGCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTCCACGACCGGGCTA 17465
QY 3962 GCGCGCAACACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 4021
Db 17464 GCGCGCAACACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 17405
QY 4022 AACCGCAGTCTCCGGGACCGCCGAATCCGCTCCGGAGCCTCAGCCCTCGGAAAGTGA 4081
Db 17404 AACCGCAGTCTCCGGGACCGCCGAATCCGCTCCGGAGCCTCAGCCCTCGGAAAGTGA 17345
QY 4082 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTGTCAGGACGATGTGAGTTCCCA 4141
Db 17344 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTGTCAGGACGATGTGAGTTCCCA 17285
QY 4142 GCCTGGGCC 4150
Db 17284 GCCTGGGCC 17276

RESULT 3

US-09-764-847-1865/c
; Sequence 1865, Application US/09764847
; Publication No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1865
; LENGTH: 17704
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1865

Query Match 10.3%; Score 429; DB 10; Length 17704;
Best Local Similarity 100.0%; Pred. No. 4.4e-104;
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3722 GTGCTGTCGCTGTCGCGCAGCATCCCGCGGCGCCCTGCTGCGGTGCGCGGACCTTCGGCCT 3781
Db 17704 GTGCTGTCGCTGTCGCGCAGCATCCCGCGGCGCCCTGCTGCGGTGCGCGGACCTTCGGCCT 17645
QY 3782 CTGCTCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 3841
Db 17644 CTGCTCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 17585
QY 3842 CGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 3901
Db 17584 CGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 17525
QY 3902 ACCGTTTGGCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTCCACGACCGGGCTA 3961
Db 17524 ACCGTTTGGCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTCCACGACCGGGCTA 17465
QY 3962 GCGCGCAACACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 4021
Db 17464 GCGCGCAACACAGCTAGCTGCAAGCGCGCGGCTCAGCGCTACCGGGGCTTCGA 17405
QY 4022 AACCGCAGTCTCCGGGACCGCCGAATCCGCTCCGGAGCCTCAGCCCTCGGAAAGTGA 4081
Db 17404 AACCGCAGTCTCCGGGACCGCCGAATCCGCTCCGGAGCCTCAGCCCTCGGAAAGTGA 17345

QY 4082 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTGTCAGGACGATGTGAGTTCCCA 4141
Db 17344 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTGTCAGGACGATGTGAGTTCCCA 17285
QY 4142 GCCTGGGCC 4150
Db 17284 GCCTGGGCC 17276

RESULT 4

US-10-032-154-1864/c
; Sequence 1864, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1864
; LENGTH: 17707
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-032-154-1864

Query Match 10.3%; Score 427.4; DB 9; Length 17707;
Best Local Similarity 99.8%; Pred. No. 1.2e-103;
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3722 GTGCTGTCGCTGTCGCGCAGCATCCCGCGGCGCCCTGCTGCGGTGCGCGGACCTTCGGCCT 3781
Db 17707 GTGCTGTCGCTGTCGCGCAGCATCCCGCGGCGCCCTGCTGCGGTGCGCGGACCTTCGGCCT 17648
QY 3782 CTGCTCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 3841
Db 17647 CTGCTCTCTCCCTCCCGCCCTTACCTCCACGCGGACCGCGCGCCAGTCAACTCCT 17588
QY 3842 CGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 3901
Db 17587 CGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACGGTC 17528
QY 3902 ACCGTTTGGCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTTCACGACCGCGGCTA 3961
Db 17527 ACCGTTTGGCAGCTCTAGCCTTTAAATTCGCGGCTCGGGACCTTCACGACCGCGGCTA 17468
QY 3962 GCGCGCAACACAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGGGCTTCGA 4021
Db 17467 GCGCGCAACACAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGGGCTTCGA 17408
QY 4022 AACCGCAGTCTCCGGGACCGCCGAATCCGCTCCGGAGCCTCAGCCCTCGGAAAGTGA 4081
Db 17407 AACCGCAGTCTCCGGGACCGCCGAATCCGCTCCGGAGCCTCAGCCCTCGGAAAGTGA 17348
QY 4082 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTGTCAGGACGATGTGAGTTCCCA 4141
Db 17347 TCCCGGCATCCGAGACCAAGATCCGCGCCCACTTGTGTCAGGACGATGTGAGTTCCCA 17288
QY 4142 GCCTGGGCC 4150
Db 17287 GCCTGGGCC 17279

RESULT 5

US-09-764-847-1864/c
; Sequence 1864, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies


```
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1864
; LENGTH: 17707
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-847-1864

Query Match          10.3%; Score 427.4; DB 10; Length 17707;
Best Local Similarity 99.8%; Pred. No. 1.2e-103;
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3722 GTGGTGTGCGTTCGCGACATCCCGCGCCCTTACCTCCAGCGGGAGCCGCGCGCTGCTGGTTCGCGGAGCTCGGCGCT 3781
Db 17707 GTGGTGTGCGTTCGCGACATCCCGCGCCCTTACCTCCAGCGGGAGCCGCGCGCTGCTGGTTCGCGGAGCTCGGCGCT 17648
QY 3782 CTGTCTCCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCGCGCGCTGCTGGTTCGCGGAGCTCGGCGCT 3841
Db 17647 CTGTCTCCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCCGCGCGCTGCTGGTTCGCGGAGCTCGGCGCT 17588
QY 3842 CGCACTTTGCCCTGCTTGGCAGCGGATTAAGGGGGCTGAGGAAATACCGGACACGGTC 3901
Db 17587 CGCACTTTGCCCTGCTTGGCAGCGGATTAAGGGGGCTGAGGAAATACCGGACACGGTC 17528
QY 3902 ACCCGTTGCAGCTCCTAGCTTTAAATTCGCGCTCGGGAGCTCCAGCACCACCGCGGCTA 3961
Db 17527 ACCCGTTGCAGCTCCTAGCTTTAAATTCGCGCTCGGGAGCTCCAGCACCACCGCGGCTA 17468
QY 3962 GCGCCGACACCACTAGCTGCAAGCGCGCGGCTCAGCGCTACCGCGGCGCTTCGA 4021
Db 17467 GCGCCGACACCACTAGCTGCAAGCGCGCGGCTCAGCGCTACCGCGGCGCTTCGA 17408
QY 4022 AACCAGCTCTCCGCGGACCGGAACTCGGCTCCGGAGGCTCAGGCCCCCTGGAAAGTGA 4081
Db 17407 AACCAGCTCTCCGCGGACCGGAACTCGGCTCCGGAGGCTCAGGCCCCCTGGAAAGTGA 17348
QY 4082 TCCCGGATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGTGAGTTCCCA 4141
Db 17347 TCCCGGATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGTGAGTTCCCA 17288
QY 4142 GCCTGGCCC 4150
Db 17287 GCCTGGCCC 17279

RESULT 6
US-09-981-353-183
; Sequence 183, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 183
; LENGTH: 5648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 464689.64
US-09-981-353-183

Query Match          9.9%; Score 409.8; DB 9; Length 5648;
Best Local Similarity 99.5%; Pred. No. 3.1e-99;
Matches 411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3719 GGTGTGTGTGTCGGTTCGGCAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3778
Db 1 GGTGTGTGTGTCGGTTCGGCAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 60
QY 3779 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 61 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 120
QY 3839 CGTCGCACCTTGGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACG 3898
Db 221 CGTCGCACCTTGGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACG 180
QY 3899 GTACACCGTTCGCGAGCTCAGCTTTAAATTCGCGCTCGGGAGCTCCACGACACCGCGG 3958
Db 281 GTACACCGTTCGCGAGCTCAGCTTTAAATTCGCGCTCGGGAGCTCCACGACACCGCGG 240
QY 3959 CTAGCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGCGGCTT 4018
Db 241 CTAGCGCGGACAAACAGCTAGCTGCAAGCGCGCGGCTCAGCGGTACCGGCGGCTT 300
QY 4019 CGAAACCGCAGTCTCTCCGCGACCCGAACTCCGCTCCCGAGGCTCAGCCCCCTGGAAG 4078
Db 301 CGAAACCGCAGTCTCTCCGCGACCCGAACTCCGCTCCCGAGGCTCAGCCCCCTGGAAG 360
QY 4079 TGATCCCGGATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGT 4131
Db 361 TGATCCCGGATCCGAGAGCAAGATGCGCGGCCACTTGTCTGAGGACGATGT 413

RESULT 7
US-09-919-172-70
; Sequence 70, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 70
; LENGTH: 5648
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 464689.64
US-09-919-172-70

Query Match          9.9%; Score 409.8; DB 10; Length 5648;
Best Local Similarity 99.5%; Pred. No. 3.1e-99;
Matches 411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3719 GGTGTGTGTGTCGGTTCGGCAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3778
Db 1 GGTGTGTGTGTCGGTTCGGCAGCATCCCGCGCCCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 60
QY 3779 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 3838
Db 61 CCTCTGTCTCTCCCTCCCGCTTACCTCCACGCGGGACCGCCCGCCAGTCAACT 120
QY 3839 CTCTCGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACG 3898
Db 121 CTCTCGCACTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGGACACG 180
QY 3899 GTACACCGTTCGCGAGCTCAGCTTTAAATTCGCGCTCGGGAGCTCCACGACACCGCGG 3958
Db 1 GTACACCGTTCGCGAGCTCAGCTTTAAATTCGCGCTCGGGAGCTCCACGACACCGCGG 3958
```

Db 181 GTACCCGTTGCCAGCTCTAGCCCTTTAAATTCCTCCGGCTCGGGACCTCCACGACCCGG 240
QY 3959 CTAGCCGCGACACACAGCTAGCTGCAAGCGCGCCGGCTCAGCGGTACCGCGGGCTT 4018
Db 241 CTAGCCGCGACACACAGCTAGCTGCAAGCGCGCCGGCTCAGCGGTACCGCGGGCTT 300
QY 4019 CGAAACCGAGTCTCTCGGCGACCCCGAACTCCGGTCCGAGGCTCAGCCCTCGAAAG 4078
Db 301 CGAAACCGAGTCTCTCGGCGACCCCGAACTCCGGTCCGAGGCTCAGCCCTCGAAAG 360
QY 4079 TGATCCCGGCATCCGAGAGCCAGATGCGCGGCCACTTCTCGAGAGCATGT 4131
Db 361 TGATCCCGGCATCCGAGAGCCAGATGCGCGGCCACTTCTCGAGAGCATAT 413
RESULT 8
US-09-764-891-7810
; Sequence 7810, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7810
; LENGTH: 13216
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7810

Query Match 8.4%; Score 349; DB 9; Length 13216;
Best Local Similarity 70.0%; Pred. No. 1,3e-82;
Matches 516; Conservative 0; Mismatches 210; Indels 11; Gaps 3;
QY 1441 TTTTCTTTTCTTTTGTGAGAAAGAGCTTGTCTTTCACCCCTGGCTGGAGGCACTGGTG 1500
Db 3625 TCTTTCTTTTCTTTTCTGGAGTCTTGTCTTTCGCGCCAGGCTGGAGTGCAGTGGTG 3684
QY 1501 TGATCTCAGCTCAGTGCACACTTTCGCTCCCGGGTTCAGCAATCTCTCTGCTCAGCCT 1560
Db 3685 CGATCTCAGCTCAGTGCACAACTCCACCTCCCGGGTTCAGAGATCTCTCTGCTTAGCCT 3744
QY 1561 CCCAAGTAGCTGGGACTACAGGCACTTCCACCATGCCAGCTGATTTTGTATTTTAG 1620
Db 3745 CCTGAGTAGCTGGGATTACAGCGGTGCACTACCGTGCCTGGCTAATTTTATATTTTAA 3804
QY 1621 TAGAGATGGGATTTACATTTGTTGGCCAGCTGGTCTCAAACTTTTGTGTCATAATTG 1680
Db 3805 TACAGATGGGTTTACCATGTTGTGTCAGGCTGGTCTCGAACTCCTGACTTCATGATCG 3864
QY 1681 TTGTAACATTTCTTCTTTGCTGAGGTAGGGCCCGCCAGCACTGTGCCAGGC --- 3920
Db 3865 TCCCCCTCGACCTCCCAAGTGTTAGGATTACAGGATGAGCCACTGTGCCAGGC --- 3920
QY 1741 AATCAAAATCAGTGTGTGGTTGACCACTGTCACTTGAGAACCACTGTGACCAAGGC 1800
Db 3921 ---GCTAATTTATTTTGTACAGACAGAGTTTCTCCATGTTGCCAGGCTGATCTGAA 3977
QY 1801 CTCAGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAGATGAATAATTTCTCCGGCC 1860
Db 3978 CTCCTGGGCGCAAGTGAACCACTCTCCCTCGGC ---CTCAAGTGTCTGGGATTAGAGCC 4034
QY 1861 AGCGTGTGGTCTATGCTGCTGAATCCAGCACTTTGGAGGCCAAGCATGTGATCAC 1920
Db 4035 GGGTGGGTGGCTCAGCGCTGTAATTCAGACACTTTGGAGGCCAAGCGAGGTGATCAG 4094
QY 1921 CTGAGGTGAGAGTTCAAACACAGCTGCGCCAACTGTGAAACCCCGTCTCTACTAAA 1980
Db 4095 CTGAGGTGAGAGTTCAGACACAGCTGCGCCAAAGGTGACACCTGTCTCTACTAAA 4154

QY 1981 ATACAAAAAATTAGTAAAGTGTGTCGCGCATGCTCTATCCAGTCTACTTGGAGGGT 2040
Db 4155 ATAC-AAATATTAGCCAGCTGTGTGGCATGCTCTATAGTCCCACTACTTGGAGGGT 4213
QY 2041 GAGCAGGAGAAATTTCTTGAACCCGGGAGCGAGGTTGCAGTGAAGCGAGATCACACCA 2100
Db 4214 GAGCAGGAGAAATCACTTGAACCCAGGAGCGGAGGTTGCAGTGAAGCGAGATCACACCA 4273
QY 2101 CTGCACTCCAGCTCGGGGAGAGGAGGAGACTTCTCTCAAAAAACAAAAACAAGA 2160
Db 4274 CTGCACTCCAGCTCGGGGAGGAGGAGTGAAGTGTCTCTCAAAAAACAAAAACAAGA 4333
QY 2161 ATTAAGCAAAATTAGACA 2177
Db 4334 ACAAAACAAACAAACA 4350
RESULT 9
US-09-764-891-7809
; Sequence 7809, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7809
; LENGTH: 29163
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7809

Query Match 8.4%; Score 347.4; DB 9; Length 29163;
Best Local Similarity 69.9%; Pred. No. 5.8e-82;
Matches 515; Conservative 0; Mismatches 211; Indels 11; Gaps 3;
QY 1441 TTTTCTTTTCTTTTGTGAGAAAGAGCTTGTCTTTCACCCCTGGCTGGAGGCACTGGTG 1500
Db 3623 TCTTTCTTTTCTTTTCTGGAGTCTTGTCTTTCGCGCCAGGCTGGAGTGCAGTGGTG 3682
QY 1501 TGATCTCAGCTCAGTGCACACTTTCGCTCCCGGGTTCAGCAATCTCTCTGCTCAGCCT 1560
Db 3683 CGATCTCAGCTCAGTGCACAACTCCACCTCCCGGGTTCAGAGATCTCTCTGCTTAGCCT 3742
QY 1561 CCCAAGTAGCTGGGACTACAGGCACTTCCACCATGCCAGCTGATTTTGTATTTTAG 1620
Db 3743 CCTGAGTAGCTGGGATTACAGCGGTGCACTACCGTGCCTGGCTAATTTTATATTTTAA 3802
QY 1621 TAGAGATGGGATTTACATTTGTTGGCCAGCTGGTCTCAAACTTTTGTGTCATAATTG 1680
Db 3803 TACAGATGGGTTTACCATGTTGTGTCAGGCTGGTCTCGAACTCCTGACTTCATGATCG 3862
QY 1681 TTGTAACATTTCTTCTTTGCTGAGGTAGGGCCCGCCAGCACTGTAATTTTCTAG 1740
Db 3863 TCCCCCTCGACCTCCCAAGTGTTAGGATTACAGGATGAGCCACTGTGCCAGGC --- 3918
QY 1741 AATCAAAATCAGTGTGTGGTTGACCACTGTCACTTGAGAACCACTGTGACCAAGGC 1800
Db 3919 ---GCTAATTTATTTTGTACAGACAGAGTTTCTCCATGTTGCCAGGCTGATCTGAA 3975
QY 1801 CTCAGAGTAGAGGTGATCTCTGCTCGAAGAGAAATAGATGAATAATTTCTCCGGCC 1860
Db 3976 CTCCTGGGCGCAAGTGAACCACTCTCCCTCGGC ---CTCAAGTGTCTGGGATTAGAGCC 4032
QY 1861 AGCGTGTGGTCTATGCTGCTGAATCCAGCACTTTGGAGGCCAAGCATGTGATCAC 1920
Db 4033 GGGTGGGTGGCTCAGCGCTGTAATTCAGCACTTTGGAGGCCAAGCGAGGTGATCAG 4092

```
QY 1921 CTGAGGTCAGGAGTCAAAACCCAGCCTGGCCAAACATGGTGAACCCCGTCTCTACTAAAA 1980
Db 4993 CTGAGGTCGCGAGTCAAGACCTGGCCCAACAAAGGTGACACCCCTGCTCTACTAAAA 4152
QY 1981 ATACAAAAAATTAGCTAAGTGTGTGGCGCATGCGTCTTAATCCAGTACTTGGGAGGT 2040
Db 4153 ATAC-AAAAATAAGCCAGGTGTGTGGCGCATGCGTCTTAATCCAGTACTTGGGAGGT 4211
QY 2041 GAGCAGAGGAATTTCTGAACCCGGGAGGAGGAGTTCAGTGAAGGAGATCACACCA 2100
Db 4212 GAGCAGAGGAATTTCTGAACCCGGGAGGAGGAGTTCAGTGAAGGAGATCACACCA 4271
QY 2101 CTGCACCTCCAGCTGGGGAGAGAGCAGACTTCTCTCAAAAAAACAACAAACAAAGA 2160
Db 4272 CTGCACCTCCAGCTGGGGAGAGAGTGAAGTGTGTCTCAAAAAAACAACAAACAAAGA 4331
QY 2161 ATTAAGCAAAATTAGACA 2177
Db 4332 ACAAAACAACAAACAA 4348

RESULT 10
US-10-020-141-7
; Sequence 7, Application US/10020141
; Publication No. US20030092013A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Ableson, Allen
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-002
; CURRENT APPLICATION NUMBER: US/10/020,141
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/313,097
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/327,485
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 26928
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-020-141-7

Query Match 8.3%; Score 342.8; DB 9; Length 26928;
Best Local Similarity 57.5%; Pred. No. 9.5e-81;
Matches 848; Conservative 0; Mismatches 552; Indels 74; Gaps 10;

QY 735 TATTTTTTTTCTTGAACAGGGTATCACTCTGTGCGCCAGGCTGGAGTACAGTGGCGTAA 794
Db 3122 TTTTITTTTTTTGGGAGTCAAGAGTCTGTCTGTGTGCGCCAGGCTGGAGTGCAGTGACACGA 3181
QY 795 TAATAGTCTACTGAGGCTCC-CCTCCTGGGCTCAAGCAATFCCGCTGGCCCTCAGCATCCT 853
Db 3182 TCTCAGTCTACTGAGGCTCCGCTCCGCGTTCAGCTATTTCTCCTCTCAGCCTCC 3241
QY 854 GAGTAGTGGGACTACAGGCTGTGGCCACAGGCGCCAGCTAAGTTTAAATAATGATTT 913
Db 3242 AAATAGTGTGAGCTATAGGCAAGCAGCATCCATGCCAGCTAATTTT-----TTAATTT 3295
QY 914 TGSTATAGAGAGCTTGTCTATGTTGCTCAGGCTGTATTTTATTTTGTGAGACAAGTGC 973
Db 3296 TAGTAGAGACGAGGTGCTCCATGGTGGCCAGGTGTGTCCTTGAAT----- 3341
QY 974 TCATGTATGTCCTATGCTCCGCTCCAGCTTCCCAAGTGTCTATCTTATCTGTCTCA 1033
Db 3342 --CTGTCTCAAGTATGATCCAGCTGCTCCGCTCCCAAGTGTGGGATGTCAGGATG 3399
QY 1034 TTAGTCAAGTACAGACATTTAGTGTGTTTCCACTTTTGTACCAATATGAAATAATCTCC 1093
Db 3400 AGACACCGCGCGCGCC-----TGCTGTGCTTCTTAAATGAGTGTCCATTTGTAA 3454
QY 1094 AGTGAATATTATGATACATTTGTGGGATATGTTTTCATTTCTGTTGGGTTTATAT 1153
```

```
Db 3455 GCTGCTGATTTCTTTGGGACATTTGTCCTGTAACCTTTTTCATAAAGCATCAGTGAATTC 3514
QY 1154 CTAGAGGTGGAAATGCTGGATCCCGGGTAATATTTTGACAGCGCAGAGTTTCAGGGAAGAA 1213
Db 3515 CCATTTCTCCACCACAGCTTCACCGTAATTTGTTGTTGTTCTTCAATTTTCAGCA 3574
QY 1214 AAATCTGGGAATAATGAAGCATGTTTAGAATATCAGCAAGAGTGCAGGGGTTTTCGGAGTT 1273
Db 3575 GAATTCATTTAGCTCTGTATAAGGCTCGCTTCAAACTGATGCTTATCCTTCTTAGTGCC 3634
QY 1274 TTAATTTATATTTCTGTTGACAAATGTC-----AGTTGATGAAGATACAAAGTTATATA 1328
Db 3635 TCAAACTACATCTCTGTTCACTCATGTTATAGCAAGTTAGTGTGATTTATTTTGTGTCAC 3694
QY 1329 AGTGAGAGTGAAGTAAGGCTGGAATAGGGCGTTCAGAGTAAATCATGAAGCACTTT 1388
Db 3695 AAAAAATTTTTTAAATCCATGCAGTCTTTTTCATAATACGCATTTTCCATGAACCTTTC 3754
QY 1339 GAATACCAAAATTAAGGAGCTTGGCTGTAAACAAATTAATAAAATCACAATTTTTTT 1448
Db 3755 GAGACCCCTTTGTAGATGCTGTTGTTTAAACCCACCCAGTTTA-----CAGTAATTTTTT 3810
QY 1449 TTTTTTTTGAGAAAGAGTCTTGTCTTTCACCTGGCTGGAGGGCAGTGGTGTGATCTCA 1508
Db 3811 TTTTTTTTGAGATGAAGTCTTGTCTGTGCGCCAGGCTGGAGTGCATTTGGCACACTCTCG 3870
QY 1509 GCTCAGTCAACTTTGCGCTCCCGGTTTCAAGCAATTCCTGCTTCAGCTCCCAAGTA 1568
Db 3871 GCTCAGTCAACTTGTGCTCTCTGCTTTCAGCAATTTTCTGCTCAGCTCCCGAGTA 3930
QY 1569 GCTGGGACTACAGCACTTCCACCATGCCAGCTGATTTTGTATTTTGTAGTAGAGATG 1628
Db 3931 GCTGGGATACAGGTTGTGCCCATGCCATGCTGCTGCTCATTAATTTTGTAGTAGAGAG 3990
QY 1629 GGATTTCACTTTGTTGGCAAGCTGCTCAAACTTTTGTGCTCATTAATTTTGTGTAAT 1688
Db 3931 GGGTTTCACTATGTTGGCTAGGCTGCTCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4050
QY 1689 ATTGTTCTTTTGTGAGGTAGGGCCCCCAGACCAAAAAAATAAATCTTGAATPCCAAA 1748
Db 4051 GGCTCCCAAGATTTGGGATTTACAGGCGTGA-----GACTCTTGCACTTGCC 4099
QY 1749 TCAGTGTGTTGTTGACCTGTGCTGAGAACCCAGTGTGACAGGCGCTCAGGAG 1808
Db 4100 TACAGTAATTTTATAGCAGCCTAGGCTAAGATAGCCATTTCTGGGTATAGAATGTCTATA 4159
QY 1809 TAGAGGTGATCTCTGCTCGAAGAGAAATAGAATGAAATATTTCTCCGGGCGCAGCGCTGG 1868
Db 4150 TACTGAACAGGCTTGCAACTGTGAGTAAGATCTGCAAGAGAGGCC-----GGGCAG 4210
QY 1859 TGGCTCATGCTGTAAATCCAGCATTTTGGAGGCGCAAGGATGTGGATACCTGAGGTC 1928
Db 4211 TGGCTCATGCTGTAAATCCAGCATTTTGGGGGCGGAGGAGGTTGATCAGCTGAGGTC 4270
QY 1929 AGGAGTTCAAAACCAAGCTGCGCAACATGGTGAACCCGCTCTCTACTAAATAACAAA 1988
Db 4271 AGCAGTTGAGACAGCCTGACCAACATGGTGAACCCCATCTCTACTAAATAACAAA 4328
QY 1939 AATTAGCTAAGTGTGGGCGCATGCTGTAATCC-----AGTACTTTG 2033
Db 4329 AATTAGCTGGGCTGGTGTGATGCTGTAATCCCTAGCATGCTTGGGAGGCTACTTG 4388
QY 2034 GGAGGCTGAGCAGGAGAAATTTCTTGAACCCGGGAGGCGAGGTTGCGAGTGAAGCGAGAT 2093
Db 4339 GGAGGCTGAGCAGGAGAAATCATTGTACTCAGAGGCGGAGGTTGCGAGTGCAGTGCAGAT 4448
QY 2094 CACACCACTGCATCCAGCCTGGGGGAGAGAGCAGAGCTTCTCTCAAAAAAACAACAAA 2153
Db 4449 CACGCCACTGCATCTCTTCTGGGTGACAGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 4508
QY 2154 CAAAAGAAATTAAGCAAAATTAGACATTTGCAGAGAG 2187
```

Db 4509 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4542

RESULT 11

US-09-880-107-2278
; Sequence 2278, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darcil T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2278
; LENGTH: 26928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M17262
US-09-880-107-2278

Query Match 8.3%; Score 342.8; DB 10; Length 26928;
Best Local Similarity 57.5%; Pred. No. 9.5e-81;
Matches 848; Conservative 0; Mismatches 552; Indels 74; Gaps 10;

QY 735 TATTTTCTTGAACAGGATATCTCTGTTCCAGGCTGGAGTACAGTGGCGTAA 794
DB 3122 TTTTCTTTGGAGTCAAGAGTCTGCTCTGTTGCCAGGCTGGAGTGCAGTGACAGA 3181

QY 795 TAATAGCTACTGCAGCTCC--CTTCCTGGGCTCAAGCAATCCGCTGGCCCTCAGCATCCT 853
DB 3182 TCTCAGCTCACTGCAGCTCCGCTCCGGGTCAAGCTATTCTCTGCTCAGGCTCC 3241

QY 854 GAGTAGCTGGGACTACAGCTGTGTCACAGGCTGTATTTTATTTGTTGAGACAAAGTC 913
DB 3242 AAATAGCTGAGACTATAGCGACATCCATGCCAGCTAATTTT-----TTATTTT 3295

QY 914 TGGTATAGAGGAGGCTTGTCTATGTTGCTCAGGCTGTATTTTATTTGTTGAGACAAAGTC 973
DB 3296 TAGTAGAGCAGAGGTGTCTCCATGTTGCCAGGTTGGTCTTGAAGT----- 3341

QY 974 TCATATGTTGCCATGATCCCGCCAGCTCCAGTCCCAAGAGTCTATCTATCTGTTCA 1033
DB 3342 --CTGCTCCTCAAGTGATCCAGCTCCCTCCGCTCCCAAGAGTGGGATTCAGGCAIG 3399

QY 1034 TTAGTCACTGACAGACATTTAGTGTGTTTCCACATTTTGTACCAATATTAATAATACCC 1093
DB 3400 AGACACCGCCCGGCC-----TGCTTGTCCCTTCTTAATAGATGTTGCTCAATTTGTA 3454

QY 1094 AGTGAATATCATGATATCATATTTGTTGGGATATGTTTTCATTTCTGTTGGGTTATAT 1153
DB 3455 GCTGCTGATTTCTTTGGGACATTTGCTCCGTAAGCTTTTCAATAAGCATCAGTGATTTCA 3514

QY 1154 CTAGAGTGGAAATGCTGGATCCCGGGTAATATTTGACAGCAGATTCAGGGGAAGAA 1213
DB 3515 CCATTTCTCCACCCAAAGTTTCAAGCTTCAATTTGTTGTTGTTCTTGTCTTCAATTTCA 3574

QY 1214 AAACCTGGGAAATCAAGCATGTTTGAATAATCAGCAAGAGTGCAGGGGTTTTCGGAGTT 1273
DB 3575 GNAATCATTTAGCTCTGATAGGGCTCCGCTCAAACTGATGCTTATCCTCTTAGTGCC 3634

QY 1274 TTTATTTTATTTCTGTTGACAAATGTC-----AGTTTGAAGATACAAAGTTATAC 1328
DB 3635 TCAAACTACATCTCTTCTCCTCATGTTATAGCAAGTTAGTGAGCTTATTTTGGTGAC 3694

QY 1329 AGTGAGAGTGAAGTAAAGGCTGGAATAGGGCTTACAGAGTAAATCATGAGCACTTT 1388
DB 3695 AAAAAATTTTAAATCCATGAGCTCTTTTTCATAATACGCACTTTCCATGACATTTTC 3754

QY 1389 GAATACCAAAATTAAGAGCTTGGCTGTAACAAAAATAATAAAAAATCAAAATTTTTTT 1448
DB 3755 GAAGACCCCTTGTAGATGCTGTTGTTTAAACACCCAGTTTA----CAGTAATTTTTT 3810

QY 1449 TTTTGTGAGAAAGAGCTTGTCTTTTCCACCTGGCTGGAGGAGGAGTGTGATCTCA 1508
DB 3811 TTTTGTGAGTGAAGCTTGTCTCTGCGCCAGGCTGGAGTGCATTCGACACATCTCG 3870

QY 1509 GCTCACTGCACTTTTCGCTCCCGGTTCAAGCAATTTCTCTGCTTACGCTCCCAAGTA 1568
DB 3871 GCTCACTGCACTCTGCTCTCTGCTTCAAGCAATTTTCTGCTCAGTCTCCCGAGTA 3930

QY 1569 GCTGGGACTACAGGCACTTCCACCATGCCAGCTGATTTTTTGTATTTTAGTAGAGTG 1628
DB 3931 GCTGGGATTACAGGTTGTGCCACCATGCTAGCTAATTTATGTTGTTTGTAGAGAG 3990

QY 1629 GATTTCACTTTTGTGGCCAGCTGCTCAAACTTTTGTGCTCATATTTGTTGTAAT 1688
DB 3991 GGGTTTCACTGTTGGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4050

QY 1689 ATTGTTCTTTTGTGAGTAGGGCCCCAGACCAAAAAATAAATCTTAGAATCCAAA 1748
DB 4051 GGCCTCCCAAGTATTGGGATTACAGGCTGA-----GACTCTTGCACTTGGCC 4099

QY 1749 TCAGTGTGTTGGTTGACCACTGCTGCTGACCAACCAAGTGTGACAGGGCTCAGAG 1808
DB 4100 TACAGTAATTTTATAGCAGCTTAGCTAGATAGCAATTTCTGGGTATAGAATCTCATA 4159

QY 1809 TAGAGTCACTCTGCTCGAAAGAAATAGAAATGAAATATTCCTCCGGCCAGGCGTGG 1868
DB 4160 TACITGAACAGGCTGCAACTGTGAGTAAAGTCTGCAAGAGGCC-----GGGCAG 4210

QY 1869 TGGCTCATGCTGTAATCCAGCACTTTGGGAGGCCAAGGCAATGATGATCACTCAGGTC 1928
DB 4211 TGGCTCATACCTGTAATCCAGCACTTTGGGGGGCGGAGGAGTGGATCACTCAGGTC 4270

QY 1929 AGGAGTTCAAAACCAAGCTGCGCAACATGGTGAACCCGCTCTCTACTAAATAACAAA 1988
DB 4271 AGCAGTTGAGACCAAGCTGACCAACATGGTGAACCCCACTCTCTACTAAATAAC--AA 4328

QY 1989 AATTAGTCAAGTGGTGGCGCATGCTGTATATCC-----AGCTACTTG 2033
DB 4329 AATTAGTGGCGTGGTAGTGCATGCTTTGTAATCCCTAGCATGCACITGGGAGCTATTG 4388

QY 2034 GGAGGTCAGGAGGAGAGATTTCTTGAACCGGGAGGAGGAGGTTGTCAGTGAAGCAGAT 2093
DB 4389 GGAGGTCAGGAGGAGAGATCACTTTGTAATCAGAGGCGGAGGTTGTCAGTGAAGAT 4448

QY 2094 CACACCACTGCACTCCAGCTGGGGAGAGAGGAGGAGTCTCTCAAAAAACAAAAA 2153
DB 4449 CAGGCCACTGCACTCTTTCTGGGTGACAGAGTGAAGTCCATCTCAAAAAACAAAA 4508

QY 2154 CAAAAGATTAAGCAATTTAGACATTTGACAGAG 2187
DB 4509 AAACAAAAACAAAAACAAAAACAAAAACCAACAG 4542

RESULT 12

US-09-764-877-3049/c
; Sequence 3049, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper

RESULT 14
US-10-031-504-1945/c
; Sequence 1945, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:

RESULT 14
US-10-031-504-1945/c
; Sequence 1945, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091.504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1945
; LENGTH: 17397
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-1945

Query Match          7.9%; Score 327.6; DB 9; Length 17397;
Best Local Similarity 57.2%; Pred. No. 8:ee-77;
Matches 1030; Conservative 0; Mismatches 569; Indels 201; Gaps 16;

QY 370 TATTATTAGACAGGTTCTCACTCTCTCAACCCAGGCTGGAGTGCAGTGGCACAATCATA 429
Db 16777 TATTATTAGACAGTTTCT-GTCCGTTGCCAGGTGCAGTGCAGTGGCACGATCTCA 16719

QY 430 GCTCACTGTAACTCAACATCCTCGG--CTCAAGTGATCTTACTACCTCACTCCACG 487
Db 16718 GCTCACTGCAACCTCTGCCCTCACTCCACCTCAGACGATCTCTCGCTCACTCCAG 16659

QY 488 TAGCTAGGACTACAGGCACACACGACCACTCTGGCTAAATTTTTTTTTTTAAATTTTCA 547
Db 16658 TAGCTGGATTACAGGCAGGTCATCACACCCGGCTAATTTTGTATT----- 16609

QY 548 TTATGTATTCATTTCTTTTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 607
Db 16608 -----TTTAGTAGAGTGGGTTCAGCATGT 16582

QY 608 TACCAGGCTGGAGGCGAGTGGCATGTGTGACAGCTGAGCAGCCCTGACTTCCTGGGCTCA 667
Db 16581 TGGCAGGCTGGT-----CTCAACTCCTGAGCTCA 16551

QY 668 AGTGATCCTCCTGCTCAGCCTCCCAAGTAGTCTGGGACTACAAACACGTGTCAACCATGCC 727
Db 16550 AGTGATCGGCTGCTCGGCCCTCCCAAGTCTGGGATTACAGGCATGAGCCACTGCGCC 16491

QY 728 TGGCTGA-----TATTTTTTTTCTTGAACAGGATACATC- TGTGCCCCAG 774
Db 16490 TGGCCAAATTATTATCATATTATTATTATTATTAGACAGAGTTTGGCTCTGTTGCCAG 16431

QY 775 GCTGGAGTACAGTGGCGCTAAATAAGTCTCACTGCAGCCTC-CCCTCCTGGGCTCAAGCAA 833
Db 16430 GCTGAGTGCAGTGGTGGGATCTGGCTCACTGCACCTCTGCTCCGGGTTCAAGGA 16371

QY 834 TCCGCTGGCTCAGCATCTCTGAGTACGTGGGACTACAGGCTGTGTGTCACAGGCCAGCT 893
Db 16370 TCTCCTGCCTCAGCCTCCCTAGTAGTGGGATTACAGCGGTGCACCACTGCTAGCT 16311

QY 894 AAGTTTAAAAAATGATTTTGGTATAGAGAGGCTTGTGCTATGTGCTCAGGCTGATTT 953
Db 16310 AATTTTGT-----ATTTTTAGTAAGACAGGGTTTCACCATGTGGCCATGCTGGTCT 16258

QY 954 TTTATTTGTGAGACAAGGCTCTCACTATGTTGCCATGATCCCCCACCCTCCACTTCCCAA 1013
Db 16257 TGAACCTCTGACCTCAGG-----TGATCCGCTGCTGCTTGGCTTCCCAA 16214

QY 1014 GTGCTCATCTTATCTGTTTCATATTAGTCAGTTGACACAGATTTTAGTGTGTTTCCACTTTTG 1073
Db 16213 GTGCTGGGATTTATAGGTGTGAGCCACCGCGCCACGCTCAAA-----TTATTATTTT 16159

QY 1074 ACCATTATGAATAATACTCCAGTGAATATTCATGTATACATTTGTGTGGGACATATGTTT 1133
Db 16158 ATTGCTGAAGAGATTTGTGCATTTTGTAGAACTTAGAAAATATAGGAGAGCAGAGAGAA 16099

QY 1134 CATTTCTGTTGGGTTTATATCTAGAGGTGGAAATGCTGGATCCCGGGTAAATATTTTGACA 1193
Db 16098 AATCAGTCATAAATTTCTCCACCC-----AAGAAATTTTCATACCCCTATAAATTTTGGCA 16043

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2003, 14:14:22 ; Search time 5373 seconds
(without alignments)
12509.077 Million cell updates/sec

Title: US-10-016-725-15
Perfect score: 4150
Sequence: 1 aggaatccatccattataat.....tgagtccacgcctgccc 4150

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estmu.*
 - 5: em_estov.*
 - 6: em_estpi.*
 - 7: em_estro.*
 - 8: em_hfc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_hfc.*
 - 12: gb_est3.*
 - 13: gb_est4.*
 - 14: gb_est5.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: gb_gss.*
 - 18: em_gss_hum.*
 - 19: em_gss_inv.*
 - 20: em_gss_pln.*
 - 21: em_gss_vrt.*
 - 22: em_gss_fun.*
 - 23: em_gss_mam.*
 - 24: em_gss_mus.*
 - 25: em_gss_other.*
 - 26: em_gss_pro.*
 - 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653.4	15.7	1238	13	BM563325
2	491.4	11.8	579	13	BI559696
3	482.4	11.6	754	13	BI827092
4	450.2	10.8	471	17	AQ760013
5	440.8	10.6	1172	13	BM546583
6	426.4	10.3	694	9	AL596692

7	422.4	10.2	469	9	AL701463
8	421.6	10.2	683	13	BI602109
9	420.8	10.1	560	9	AL708245
10	407.4	9.8	1012	13	BM465118
11	406.2	9.8	439	9	AL701687
12	404.4	9.7	1024	13	BM451306
13	402.8	9.7	871	14	BQ431921
14	402.4	9.7	983	13	BM480208
15	400.4	9.6	546	9	AL556514
16	364.8	8.8	1015	13	BM476059
17	353.6	8.5	733	13	BI545552
18	329.8	7.9	367	9	AL589552
19	328.4	7.9	785	9	AU130094
20	299.4	7.2	683	9	AU123589
21	279.6	6.7	300	9	AU098830
22	279.6	6.7	309	9	AU076658
23	255.8	6.2	874	13	BM051176
24	255.8	6.2	2971	17	AF101960
25	249.4	6.0	524	10	AW973992
26	249.2	6.0	580	14	BQ053726
27	248.2	6.0	731	9	AI687343
28	247.2	6.0	775	14	BM905333
29	246.8	5.9	598	14	BQ017808
30	246.2	5.9	508	17	AQ418545
31	245.4	5.9	467	17	AQ388230
32	245.4	5.9	785	14	BM721013
33	244.8	5.9	663	17	AG174449
34	244.6	5.9	750	14	BM678642
35	243.8	5.9	2971	17	AF101960
36	242.6	5.8	493	10	AW328331
37	242.4	5.8	532	12	BG830539
38	242.4	5.8	861	12	BG676827
39	242.2	5.8	759	14	BQ436286
40	241.8	5.8	1750	11	AF119908
41	241.6	5.8	717	17	AG080871
42	241.6	5.8	967	12	BE897079
43	241.4	5.8	585	13	BM045436
44	240.8	5.8	535	17	AQ395337
45	240.8	5.8	797	12	BG676969

ALIGNMENTS

RESULT 1
BM563325
LOCUS 1238 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6564864 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5743936
5', mRNA sequence.
ACCESSION BM563325
VERSION BM563325.1 GI:18810153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1238)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM12764 row: h column: 17
High quality sequence stop: 622.
Location/Qualifiers
1. .1238

FEATURES
source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5743936"
 /clone_lib="NIH_MGC_119"
 /tissue_type="medulla"
 /lab_host="DH108"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
 BASE COUNT 224 a 418 c 370 g 226 t
 ORIGIN

Query Match 15.7%; Score 653.4; DB 13; Length 1238;
 Best Local Similarity 98.6%; Pred. No. 1.2e-66;
 Matches 701; Conservative 0; Mismatches 6; Indels 4; Gaps 4;
 QY 3266 TGGGTGGCTGCCACGCTCCCTGCCACGCGCTGGGGGACGGAGACGGGACGAGA 3325
 Db 20 TGGGTGGCTGCCACGCTCCCTGCC-CGCGCTGGGGGACGGGAGACGGGACGAGA 78
 QY 3326 TGTAGTGGTGGCGCCCGCCGAGGTTACCACTGTTCTTGAGAACTTCCCGAGTGC 3385
 Db 79 TGTAGTGGTGGCGCCCGCCGAGGTTACCACTGTTCTTGAGAACTTCCCGAGTGC 138
 QY 3386 CCACCCACCGCTTCCTGGTGGCCGAGGCGCGTCTGGCTAGGCTCCGCGCCCGACG 3445
 Db 139 CCACCCACCGCTTCCTGGTGGCCGAGGCGCGTCTGGCTAGGCTCCGCGCCCGACG 198
 QY 3446 CCACAAACCGGTCCCGAGCCCTTCACAGAGAAAGCTCCGACGGGATGCCGGGACG 3505
 Db 199 CCACAAACCGGTCCCGAGCCCTTCACAGAGAAAGCTCCGACGGGATGCCGGGACG 258
 QY 3506 AGCCACAGCGCGGTGGAGAGAGAGTGAAGAGGAGAAACAGAGGGGAGGGGAGCGAG 3565
 Db 259 AGCCACAGCGCGGTGGAGAGAGAGTGAAGAGGAGAAACAGAGGGGAGGGGAGCGAG 318
 QY 3566 GAGTGGCGCAGAGGAGAAACAGAGATGGCGCGAGCCATGCAACGGCAGGACGAGGT 3625
 Db 319 GAGTGGCGCAGAGGAGAAACAGAGATGGCGCGAGCCATGCAACGGCAGGACGAGGT 378
 QY 3626 GGCACCAATTCCTTCGGCCATGACGAGCGGAGTTTACAGAAGCTCATTAGCATTT 3685
 Db 379 GGCACCAATTCCTTCGGCCATGACGAGCGGAGTTTACAGAAGCTCATTAGCATTT 438
 QY 3686 CCCACAGGCGAGGCGAGGCGCAGAGCGGGTGGTGTGGTGTGCGGTGTCGGCAGCATCC 3745
 Db 439 CCCACAGGCGAGGCGAGGCGCAGAGCGGGTGGTGTGGTGTGCGGTGTCGGCAGCATCC 498
 QY 3746 CCGSGCCCTGCTGCGTGCCTCCGAGCCCTCGCCCTGCTGCT-CCCGCTCCCGCCCT 3804
 Db 499 CCGSGCCCTGCTGCGTGCCTCCGAGCCCTCGCCCTGCTGCT-CCCGCTCCCGCCCT 558
 QY 3805 TACTCCACCGGAGACCGCCGCGCCAGTCACTCCCTCGCACTTTGCGCTGCTTGGCAG 3864
 Db 559 TACTCCACCGGAGACCGCCGCGCCAGTCACTCCCTCGCACTTTGCGCTGCTTGGCAG 618
 QY 3865 CGGATAAAGGGGCTGAGAAATACCGGACAGGTCACCGCTTGGCAGCTTAGGCTTT 3924
 Db 619 CGGATAAAGGGGCTGAGAAATACCGGACAGGTCACCGCTTGGCAGCTTAGGCTTT 678
 QY 3925 AATTCCTCCGTCGGGACTTCACGACAC-GGGGTACGCGCGGACAC 3973
 Db 679 TAATTCCTCCGCTCGGGAGCTTCCCGACCGGGGGCTAGCGCGGACAC 729

RESULT 2
 BI559696

LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BI559696
 60325256FF1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5294921 5',
 mRNA sequence.
 BI559696
 BI559696.1 GI:15447010
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 579)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11746 row: c column: 18
 High quality sequence stop: 576.

FEATURES
 source

Location/Qualifiers
 1..579
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5294921"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH108"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
 size selected for average insert size 2.2 kb and
 normalized to 500 ng. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 103 a 188 c 205 g 82 t 1 others

Query Match
 Best Local Similarity 97.6%; Pred. No. 8.4e-48;
 Matches 562; Conservative 0; Mismatches 7; Indels 7; Gaps 6;

QY 3229 GGAGCGGGGCTTCCCGCGTCCCAAGCTCCAGATCCTGGGGTGGCTGCCACGTCCTCCT 3288
 Db 4 GGAGCGGGGCTTCCCGCGTCCCAAGCTCCAGATCCTGGGGTGGCTGCCACGTCCTCCT 63
 QY 3289 GCCACGGCTTGGGGGACGGGAGACGGGACGGAGATGTTAGTGGTGGGCGCCCCCG 3347
 Db 64 GCCACGGCTTGGGGGACGGGAGACGGGACGGAGATGTTAGTGGTGGGCGCCCCCG 123
 QY 3348 AGGG--TTTACCCTGTTTCTTGAGAACTTCCAGTCCACCCACCCGCTTCTCGTG 3405
 Db 124 AGGGTTCACNCACTGTTTCTTGAGAACTTCCAGTCCACCCACCCGCTTCTCGTG 193
 QY 3406 TGCCCGAGGGGCTTCTGGGCTAGGCTCCGCGCCCGCCAGCCCAACCCGCGTCCAGCC 3465
 Db 184 TGCCCGAGGGGCTTCTGGGCTAGGCTCCGCGCCCGCCAGCCCAACCCGCGTCCAGCC 243
 QY 3466 CTTCCAGAGAGAAAGTCCCGACGGGATGCCGGGACGAGGCCCGGGCGGGTGGAA 3525
 Db 244 CTTCCAGAGAGAAAGTCCCGACGGGATGCCGGGACGAGGCCCGGGCGGGTGGAA 303
 QY 3526 GAGAAGCTGAGAAGAGAAACAGAGGGAGGG--GGAGCGAGGAGCTGGCGGACAGAGGAA 3584
 Db 304 GAGAAGCTGAGAAGAGAAACAGAGGGAGGGCGGAGGAGCTGGCGGACAGAGGAA 363

QY	3585	CAGCAGATTGGCGGAGGCCAATGSCAA-CGGCAGGACGAGGTGCACAAATTCCTTCG	3643
Db	364	CAGCAGATTGGCGGAGGCCAATGSCAAGCGGACGAGGTGGCACCARAATTCCTTCG	423
QY	3644	GCCAAATCAGCAGCGGGAGTTTACAGAAGCCTCATTAGCATTTCCCCAGAGGC-AGGGGCA	3702
Db	424	GCCAAATCAGCAGCGGGAGTTTACAGAAGCCTCATTAGCATTTCCCCAGAGGTAGGGGCA	483
QY	3703	GGGCGAGAGCGCGGGTGGTGTGTCGGTGTCGGCAGCATCCCGCGCCCTGCTGCGG	3762
Db	484	CGGCGAGAGCGCGGGTGGTGTGTCGGTGTCGGCAGCATCCCGCGCCCTGCTGCGG	543
QY	3763	TGCGCGGAGCGCTCGGCGCTGTGCTCTCCTCCGCCCTCC	3798
Db	544	TGCGCGGAGCGCTCGG-CTCTGTCTCTCCGCCCTCC	578
RESULT 3			
BI827092			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
Source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			
QY			
Db			

FEATURES


```

mRNA sequence.
ACCESSION B1602109
VERSION B1602109.1 GI:15495035
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 683)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1A11729 row: p column: 04
High quality sequence stop: 675.
FEATURES
source
1..683
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5288691"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NIH), National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 140 a 238 c 179 g 126 t
ORIGIN
Query Match 10.2%; Score 421.6; DB 13; Length 683;
Best Local Similarity 99.1%; Pred. No. 7.9e-40;
Matches 424; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3704 GGGCAGAGCGGGTGTGTGGTGTGGTGTGGCAGCATCCCGCGCGCTGCTGCGGT 3763
Db 2 GCGGAGAGCGGGTGTGTGGTGTGGTGTGGCAGCATCCCGCGCGCTGCTGCGGT 61

QY 3764 CGCCGCGAGCGCTCGGCTCTGTCTCTCCCTCCCGCTTACCTCCAGCGGGACCGC 3823
Db 62 CGCGCGAGCGCTCGGCTCTGTCTCTCTCCCTCCCGCTTACCTCCAGCGGGACCGC 121

QY 3824 CGCGCGCAGTCACTCTCGCAGTTTCGCCGTGTGGCAGCGGATAAAGGGGGCTGAG 3883
Db 122 CCAGCGCAGTCACTCTCGCAGTTTCGCCGTGTGGCAGCGGATAAAGGGGGCTGAG 181

QY 3884 GAATACCGGACGGTCAACCGTTGCGAGTCTAGCTTTAAATTCGCGCTCGGGAC 3943
Db 182 GAATACCGGACGGTCAACCGTTGCGAGTCTAGCTTTAAATTCGCGCTCGGGAC 241

QY 3944 CTCACGCGACCGGGCTAGCGCGGACACACAGCTAGCGTGAAGCGCGCGGCTCAGCG 4003
Db 242 CTCACGCGACCGGGCTAGCGCGGACACACAGCTAGCGTGAAGCGCGCGGCTCAGCG 301

QY 4004 CGTACCGCGGGCTCGAAGCGGAGTCTCGCGGACCGGAACTCCGCTCCGGAGCCT 4063
Db 302 CGTACCGCGGGCTCGAAGCGGAGTCTCGCGGACCGGAACTCCGCTCCGGAGCCT 361

QY 4064 CAGCCCCCTGGAAGTATCCCGCATCCGAGAGCGCAAGATGCCGGGCCACTTGTCTGAG 4123

```

```

|||||
Db 362 CAGCCCCCTGGAAGTATCCCGCATCCGAGAGCCAGATCCCGCCACTTGTCTGAG 421
QY 4124 GACGATCT 4131
Db 422 GACGATAT 429
|||||

RESULT 9
AL708245 560 bp mRNA linear EST 22-MAR-2002
LOCUS DKFZp686P0950.r1.686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKFZp686P0950 5', mRNA sequence.
ACCESSION AL708245
VERSION AL708245.1 GI:19691600
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 560)
Ottawaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Well, B. and
Wiemann, S.
EST (Ottawaelder, B., Obermaier, B., Mewes, H.W., Well, B. and Wiemann
S.)
Unpublished (2001)
Contact: Ottawaelder B
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medigenomix (Martinsried/Germany) within the CDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKFZp686P0950) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1..560
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686P0950"
/clone_lib="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 107 a 202 c 154 g 97 t
ORIGIN
Query Match 10.1%; Score 420.8; DB 9; Length 560;
Best Local Similarity 99.5%; Pred. No. 1.1e-39;
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3708 AGAGCGCGGGTGTGGTGTGGTGTGGCAGCATCCCGCGCGCTGCTGCGGTGCC 3767
Db 1 AGAGCGCGGGTGTGGTGTGGTGTGGCAGCATCCCGCGCGCTGCTGCGGTGCC 60

QY 3768 GCGAGCTCGGCGCTGTCTCTCCCTCCCGCTTACCTCCAGCGGACCGCGCGC 3827
Db 61 GCGAGCTCGGCGCTGTCTCTCCCTCCCGCTTACCTCCAGCGGACCGCGCGC 120

QY 3828 GCCAGTCACTCTCGCAGCTTTGCCCTTGTGGCAGCGGATAAAGGGGGCTGAGAAA 3887
Db 121 GCCAGTCACTCTCGCAGCTTTGCCCTTGTGGCAGCGGATAAAGGGGGCTGAGAAA 180

QY 3888 TACCGGACAGGTCACCGGTGCCAGCTCTAGCCCTTTAAATTCGCGCTCGGGACCTCC 3947
Db 181 TACCGGACAGGTCACCGGTGCCAGCTCTAGCCCTTTAAATTCGCGCTCGGGACCTCC 240

QY 3948 ACGCAGCGGGCTAGCGCGGACACAGCTAGCTAGCGTGAAGGCGCGCGGCTCAGCGGTA 4007

```

```

Db 241 ACGCACCGCGCTAGCGCGGACACACAGCTAGCGTGCAGCGCGCGCTCAGCGCGTA 300
|||||
QY 4008 CCGCGGGCTTCGAACCGCAGTCTCCGGGAGACCCCAACTCCGCTCCGGAGCCTCAGC 4067
|||||
Db 301 CCGCGGGCTTCGAACCGCAGTCTCCGGGAGACCCCAACTCCGCTCCGGAGCCTCAGC 360
|||||
QY 4068 CCGCTGGAAGTATCCCGGCATCCGAGACCAAGATCCGCGCCACATTCGTCAGGACG 4127
|||||
Db 361 CCGCTGGAAGTATCCCGGCATCCGAGACCAAGATCCGCGCCACATTCGTCAGGACG 420
|||||
QY 4128 ATGT 4131
|||
Db 421 ATAT 424
|||

RESULT 10
BM465118 1012 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6427715 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5501839
5', mRNA sequence.
ACCESSION BM465118
VERSION BM465118.1 GI:18514160
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1012)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12139 row: a column: 08
High quality sequence stop: 568.
location/Qualifiers
1. 1012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5501839"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 203 a 350 c 255 g 204 t
ORIGIN
Query Match 9.8%; Score 407.4; DB 13; Length 1012;
Best Local Similarity 99.8%; Pred. No. 2.5e-38;
Matches 408; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3723 TGGTGTGCGTGTGCGCAGCATCCCGGGCCCTGCTGCGGTGCGCGCGAGCCTCGGCCTC 3782
|||||
Db 1 TGGTGTGCGTGTGCGCAGCATCCCGGGCCCTGCTGCGGTGCGCGCGAGCCTCGGCCTC 60
|||||
QY 3783 TGTCTCTCCCTCCCTCCCTTACTCTCCACGGGACCGCCGCGCCAGTCAACTCCTC 3842
|||||
Db 61 TGTCTCTCCCTCCCTCCCTTACTCTCCACGGGACCGCCGCGCCAGTCAACTCCTC 120
|||||
QY 3843 GCACATTGCGCTGTGGCAGCGGATTAAGGGGGCTGAGGAATACGGACAGGCTCA 3902
|||||
Db 121 GCACATTGCGCTGTGGCAGCGGATTAAGGGGGCTGAGGAATACGGACAGGCTCA 180
|||||

```

```

QY 3903 CCCGTCGACGCTAGCCTTTAAATCCCGGCTCGGGACCTCCACGCAACCGCGCTAG 3962
|||||
Db 181 CCGGTGGCAGCTTAGCCTTTAAATCCCGGCTCGGGACCTCCACGCAACCGCGCTAG 240
|||||
QY 3963 CGCCGACAAACAGCTAGCGTGCAAGCGCGCGCGGTCTACGCGGTACCGCGCGGCTTCGAA 4022
|||||
Db 241 CGCCGACAAACAGCTAGCGTGCAAGCGCGCGCGGTCTACGCGGTACCGCGCGGCTTCGAA 300
|||||
QY 4023 ACCGCACTCTCCCGGAGACCGGAACTCGGCTCCGGAGCCTCAGCCCTGGAAGTAT 4082
|||||
Db 301 ACCGCACTCTCCCGGAGACCGGAACTCGGCTCCGGAGCCTCAGCCCTGGAAGTAT 360
|||||
QY 4083 CCGGCATCCGAGAGCAAGATCCGCGGCCACTTGTGTCAGGACGATGT 4131
|||||
Db 361 CCGGCATCCGAGAGCAAGATCCGCGGCCACTTGTGTCAGGACGATAT 409
|||||

RESULT 11
AL701687 439 bp mRNA linear EST 22-MAR-2002
LOCUS DKEZp686807148_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKEZp686807148 5', mRNA sequence.
ACCESSION AL701687
VERSION AL701687.1 GI:19685043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 439)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
,S.
EST (Duesterhoeft, et al.)
TITLE Unpublished (1999)
JOURNAL Contact: Duesterhoeft A
COMMENT MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No si sequence available.
This clone (DKFZp686807148) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
location/Qualifiers
1. 439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKEZp686807148"
/clone_lib="586 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI;
cDNA collection"
BASE COUNT 69 a 169 c 129 g 72 t
ORIGIN
Query Match 9.8%; Score 406.2; DB 9; Length 439;
Best Local Similarity 99.1%; Pred. No. 6.4e-38;
Matches 419; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 3709 GAGCGCGGTGTGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 3768
|||||
Db 2 GAGCGCGGTGTGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 61
|||||
QY 3769 CGAGCTCGGCTCTGTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 3828
|||||
Db 62 CGAGCTCGGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 121
|||||

```

FEATURES

source


```

QY 3829 CCAGTCAACTCTCGACCTTGGCCCTTGGTGGAGCGAGTAAAGAGGGGGCTGAGAAAT 3888
Db 122 CCAGTCAACTCTCTCGACCTTGGCCCTTGGTGGAGCGGAT-AAAGGGGGCTGAGAAAT 180
QY 3889 ACCGACACGGGTACCGCTTGGCAGCTTAGCTTTAAATCCCGGCTCGGGAGCTCCA 3948
Db 181 ACCGACACGGGTACCGCTTGGCAGCTTAGCTTTAAATCCCGGCTCGGGAGCTCCA 240
QY 3949 CGCACCGGGCTAGCGCGGACAAACAGCTAGCGTCAAGGGCGCGGCTCAGCGCGTAC 4008
Db 241 CGCACCGGGCTAGCGCGGACAAACAGCTAGCGTCAAGGGCGCGGCTCAGCGCGTAC 300
QY 4009 CGCGGGCTTCAACCGGAGTCTCCGGGACCCCGAAGCTCCGCTCGGGAGCTCAGCC 4068
Db 301 CGCGGGCTTCAACCGGAGTCTCCGGGACCCCGAAGCTCCGCTCGGGAGCTCAGCC 360
QY 4069 CCTCGGAAGTATCCCGGATCCGAGAGCCAAAGATCCGGCCACTTGTCTCAGGAGA 4128
Db 361 CCTCGGAAGTATCCCGGATCCGAGAGCCAAAGATCCGGCCACTTGTCTCAGGAGA 420
QY 4129 TGT 4131
Db 421 TAT 423

RESULT 12
BM451306 1024 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6386406 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493817
DEFINITION 5', mRNA sequence.
ACCESSION BM451306
VERSION BM451306.1 GI:18500346
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1024)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM2118 row: c column: 02
High quality sequence stop: 564.
FEATURES
source
Location/Qualifiers
1..1024
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5493817"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 219 a 338 c 252 g 214 t 1 others
ORIGIN
Query Match 9.78; Score 404.4; DB 13; Length 1024;
Best Local Similarity 99.8%; Pred. No. 5.6e-38;
Matches 405; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3726 TGTCGGTGTGGCAGCATCCCGCGGCGCTGCTCGGTGCGCGAGCGCTGCTGT 3785
|||||

```

```

Db 1 TGTGGTGTGGCAGCATCCCGCGGCGCTGCTGCGGTGCGCGAGCGCTCGGCTCTGT 60
QY 3786 CTCCTCCCGCTCCCGCTTACTTCCACGCGGAGCGCGCGCCAGTCAACTCTCGCA 3845
Db 61 CTCCTCCCGCTCCCGCTTACTTCCACGCGGAGCGCGCGCCAGTCAACTCTCGCA 120
QY 3846 CTTTCCCGCTTGTGGCAGCGGATAAAGGGGGCTGAGAAATACGGACACGGTCACCC 3905
Db 121 CTTTCCCGCTTGTGGCAGCGGATAAAGGGGGCTGAGAAATACGGACACGGTCACCC 180
QY 3906 GTTGCAGCTTAGCTTTAAATCCCGGCTCGGGAGCTCCACGACCGCGGCTAGGCG 3965
Db 181 GTTGCAGCTTAGCTTTAAATCCCGGCTCGGGAGCTCCACGACCGCGGCTAGGCG 240
QY 3966 GCACAACAGCTAGCTGCAAGGCGCGGCTCAGCGGTACGGCGGGCTTCGAAACC 4025
Db 241 GCACAACAGCTAGCTGCAAGGCGCGGCTCAGCGGTACGGCGGGCTTCGAAACC 300
QY 4026 GCAGTCTCCGCGGACCCCGAAGTCCGCTCCGAGGCGCTCAGCGGCTGAAAGTATGCC 4085
Db 301 GCAGTCTCCGCGGACCCCGAAGTCCGCTCCGAGGCGCTCAGCGGCTGAAAGTATGCC 360
QY 4086 GGCATCCGAGAGCCAAAGATCCGGCGGCGCTGCTCAGGAGCATGT 4131
Db 361 GGCATCCGAGAGCCAAAGATCCGGCGGCGCTGCTCAGGAGCATGT 406

RESULT 13
BM431921 871 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7826820 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6153075
DEFINITION 5', mRNA sequence.
ACCESSION BM431921
VERSION BM431921.1 GI:21170997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 871)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13491 row: p column: 04
High quality sequence stop: 565.
FEATURES
source
Location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6153075"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 183 a 294 c 218 g 176 t
ORIGIN
Query Match 9.78; Score 402.8; DB 14; Length 871;
Best Local Similarity 99.5%; Pred. No. 9.5e-38;
Matches 404; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3726 TGTCGGTGTGCGACGATCCCGGGGCGCTGCTGCGGTGCGCGGAGCCTCGGCCTCTGT 3785
Db 1 TGTGCGTGTAGCGACATCCCGGGCGCTGCTGCGGTGCGCGGAGCCTCGGCCTCTGT 60

QY 3786 CTCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCGCCGCGGAGTCACTCCTCCGA 3845
Db 61 CTCCTCCCGCTCCCGCCCTTACCTCCAGCGGGAGCGCCGCGGAGTCACTCCTCCGA 120

QY 3846 CTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGACACGGTCAACC 3905
Db 121 CTTTGCCCTGCTTGGCAGCGGATAAAAGGGGCTGAGGAAATACCGACACGGTCAACC 180

QY 3906 GTTGCCAGCTTAGCTTTAAATCCCGGCTCGGGGAGCTCCAGCACCGCGGTAGCGC 3965
Db 181 GTTGCCAGCTTAGCTTTAAATCCCGGCTCGGGGAGCTCCAGCACCGCGGTAGCGC 240

QY 3966 CGACAACAGCTAGCTGTCAGAGCGCGCGGCTCAGCGCGTACCGCGGGCTTCGAAC 4025
Db 241 CGACAACAGCTAGCTGTCAGAGCGCGCGGCTCAGCGCGTACCGCGGGCTTCGAAC 300

QY 4026 GCAGTCTCTCCGCGACCCCGAACTCCGCTCCGGAGCCTCAGCCCTCGAAAGTGATCCC 4085
Db 301 GCAGTCTCTCCGCGACCCCGAACTCCGCTCCGGAGCCTCAGCCCTCGAAAGTGATCCC 360

QY 4086 GGCATCCGAGAGCCCAAGATGCCGCGCCACTTGTCTCAGGACGATGT 4131
Db 361 GGCATCCGAGAGCCCAAGATGCCGCGCCACTTGTCTCAGGACGATAT 406

RESULT 14
BM480208
LOCUS BM480208
DEFINITION AGENCOURT_6468359 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574129
5', mRNA sequence.
ACCESSION BM480208
VERSION BM480208.1 GI:18529250
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12322 row: e column: 10
High quality sequence stop: 659.
FEATURES
source
1..983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 207 a 330 c 242 g 204 t
ORIGIN

Query Match 9.7%; Score 402.4; DB 13; Length 983;

```

```

Best Local Similarity 98.5%; Pred. No. 9.7e-38;
Matches 406; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3720 GTGTGGTGTGCGGTGTCGGAGCATCCCGGGCGCCCTGCTGCGGTGCGCGGAGCCTCGGC 3779
Db 6 GCGTCCGGTGTGCGGTGTCGGAGCATCCCGGGCGCCCTGCTGCGGTGCGCGGAGCCTCGGC 65

QY 3780 CTCGTCTCCCTCCCGCTCCCGGCTTACTCCACGGGACCGCGCGCCAGTCAATC 3839
Db 66 CTCGTCTCCCTCCCGCTCCCGGCTTACTCCACGGGACCGCGCGCCAGTCAATC 125

QY 3840 CTCGCACCTTTCGCCCTGCTTGCACGGATAAAGGGGCTGAGGAAATACCGGACACGG 3899
Db 126 CTCGCACCTTTCGCCCTGCTTGCACGGATAAAGGGGCTGAGGAAATACCGGACACGG 185

QY 3900 TCACCGCTTGCAGCTCTAGCTTTAAATCCCGGCTCGGGGACCTCCACGACCGCGGC 3959
Db 186 TCACCGCTTGCAGCTCTAGCTTTAAATCCCGGCTCGGGGACCTCCACGACCGCGGC 245

QY 3960 TAGCGCCGACACCACTAGCTGCAAGCGCGCGGCTCAGCGGTACCGCGGGCTTC 4019
Db 246 TAGCGCCGACACCACTAGCTGCAAGCGCGCGGCTCAGCGGTACCGCGGGCTTC 305

QY 4020 GAAACCGCAGTCTCTCCGCGACCCCGAACTCCGCTCCGAGCCTCAGCCCTCGAAAGT 4079
Db 306 GAAACCGCAGTCTCTCCGCGACCCCGAACTCCGCTCCGAGCCTCAGCCCTCGAAAGT 365

QY 4080 GATCCCGCATCCGAGAGCCACAGATGCCGCGCCACTTGTCTCAGGACGATGT 4131
Db 366 GATCCCGCATCCGAGAGCCACAGATGCCGCGCCACTTGTCTCAGGACGATAT 417

RESULT 15
AL556514
LOCUS AL556514
DEFINITION AL556514 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK006Y012 5
prime, mRNA sequence.
ACCESSION AL556514
VERSION AL556514.1 GI:12899258
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 546)
Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..546
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK006Y012"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="Placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 97 a 195 c 146 g 95 t 13 others
ORIGIN

Query Match 9.6%; Score 400.4; DB 9; Length 546;

```

Best Local Similarity 97.5%; Pred. No. 2.5e-37;
Matches 398; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 3724 GGTGTCGGTGTGGCAGCATCCCGGCGCCCTTGTGTCGGTTCGCGCAGCCTCGGCCCTCT 3783
Db 1 GGTGTCGGTGTGGCAGCATCCCGGCGCCCTTGTGTCGGTTCGCGCAGCCTCGGCCCTCT 60

QY 3784 GTCTCTCCCTCCCTCCCGCCCTTACCTCCACGGGACCGCCCGGCCAGTCAACTCCTCG 3843
Db 61 GTCTCTCTCCCTCCCGCCCTTACCTCCACGGGACCGCCCGGCCAGTCAACTCCTCG 120

QY 3844 CACTTTGCCCTCTGTGGCAGCGGATAAAAGGGGCTGAGGAATACCGGACACGGTCAC 3903
Db 121 CACTTTGCCCTCTGTGGCAGCGGATAAAAGGGGCTGAGGAATACCGGACACGGTCAC 180

QY 3904 CCGTTGCCAGCTCTAGCCTTTAAATTCCTGGGACCTCCACGACCGCGGCTAGC 3963
Db 181 CCGTTGCCAGCTCTAGCCTTTAAATTCCTGGGACCTCCACGACCGCGGCTAGC 240

QY 3964 GCCGACACCAAGCTAGCTGCAAGCGCGCGCTCAGCGGTACCGGGGCTTCGAAA 4023
Db 241 GCCGACACCAAGCTAGCTGCAAGCGCGCGCTCAGCGGTACCGGGGCTTCGAAA 300

QY 4024 CCGCAGTCTCTCCGGCAGCCCGGAATCCGCTCCGGAGCCTCAGCCCTCGAAAAGTGATC 4083
Db 301 CCGCAGTCTCTCCGGCAGCCCGGAATCCGCTCCGGAGCCTCAGCCCTCGAAAAGTGATC 360

QY 4084 CCGGCATCCGAGAGCCAGATCGCGGCCACTTGTGTCAGGACGATGT 4131
Db 361 CCGGCMTGGAGAGCCMAGATCGCGGCCACTTGTGTCAGGACGATAT 408

Search completed: May 20, 2003, 19:06:17
Job time : 5452 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:17:20 ; Search time 45 Seconds
(without alignments)
47.378 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92

Sequence: 1 MPAHLQDDVSPFAPW 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	92	100.0	16	ABB76307	Human stearyl-CoA
2	56	60.9	359	AAV69378	Amino acid sequenc
3	56	60.9	359	ABB44583	Human wound healin
4	49	53.3	285	ABG05270	Novel human diagno
5	49	53.3	285	ABG26777	Novel human diagno
6	47	51.1	334	ABG06715	Novel human diagno
7	47	51.1	334	ABG26818	Novel human diagno
8	45	48.9	1675	AAU75109	Clathrin heavy cha
9	44.5	48.4	61	ABP07968	Human ORFX protein
10	43.5	47.3	518	ABB92812	Herbicidally activ

11	43.5	47.3	632	23	AAE20083	Lactobacillus rham
12	43	46.7	437	23	ABP26786	Streptococcus poly
13	42	45.7	96	23	ABP33456	Human ORF2429 prot
14	42	45.7	184	21	AAV70255	Canine angiotensin
15	42	45.7	184	23	AAO17430	Canine endostatin.
16	42	45.7	230	23	AAO17429	Canine pro-endosta
17	42	45.7	255	21	AAV70405	Class 1 fatty acid
18	42	45.7	423	22	ABG06816	Novel human diagno
19	42	45.7	423	22	ABG29289	Novel human diagno
20	41.5	45.1	388	21	ABB32493	S. lavendulae Mit
21	41	44.6	114	23	ABP09080	Human ORFX protei
22	41	44.6	237	22	ABG05288	Novel human diagno
23	41	44.6	423	22	ABG26779	Novel human diagno
24	41	44.6	1146	23	ABB91737	Herbicidally activ
25	40.5	44.0	51	22	AAW85247	Human immune/haema
26	40.5	44.0	257	21	ABG28917	Arabidopsis thalia
27	40.5	44.0	328	21	ABG28916	Arabidopsis thalia
28	40.5	44.0	431	12	AAK14617	Encoded by (interr
29	40	43.5	81	22	AAAG2815	C glutamicum prote
30	40	43.5	128	22	ABB10756	Human ovarian and/
31	40	43.5	128	22	AAW94131	Human reproductive
32	40	43.5	197	14	AAK37919	Cyn di derived fro
33	40	43.5	197	14	AAK37920	Cyn di derived fro
34	40	43.5	200	14	AAK37913	Cyn di derived fro
35	40	43.5	243	8	AAW70483	Sequence encoded b
36	40	43.5	253	14	AAK37921	Cyn di derived fro
37	40	43.5	255	21	AAV70406	Class 2 fatty acid
38	40	43.5	272	14	AAK37917	Cyn di.18 Cynodo
39	39.5	42.9	72	22	ABB43771	Peptide #11277 enc
40	39.5	42.9	72	22	ABB26714	Protein #8713 enco
41	39.5	42.9	72	22	AAW64733	Human brain expres
42	39.5	42.9	72	22	AAW77503	Human bone marrow
43	39.5	42.9	72	22	AAW21433	Peptide #7867 enco
44	39.5	42.9	72	22	AAK37675	Peptide #11712 enc
45	39.5	42.9	72	23	ABG46528	Human peptide enco

ALIGNMENTS

RESULT 1
ABB76307
ID ABB76307 standard; Peptide; 16 AA.
XX AC ABB76307;
XX AC
XX DT 22-AUG-2002 (first entry)
XX DE Hunan stearyl-CoA desaturase N-terminal peptide..
XX DE
XX KW Stearyl-CoA desaturase; SCD; enzyme; human; promoter; virucide;
KW dermatological; cytostatic; immunosuppressive; antiallergic;
KW antiarthritic; antiinflammatory; cardiovascular; antianaemic;
KW gene therapy.
XX OS Homo sapiens.
XX PN WC200236780-A2.
XX PD 10-MAY-2002.
XX PF 31-OCT-2001; 2001WO-US45199.
XX PR 31-OCT-2000; 2000US-244508P.
XX PR 30-OCT-2001; 2001US-0244508.
XX PA (JOHU) JOHNSON & JOHNSON CONSUMER CO INC.
XX PI Prouty SM, Zhang L, Steen KS;
XX WP; 2002-471502/50.
XX DR N-PSDB; ABL57445.

PT New human stearoyl-CoA desaturase gene promoter, useful for treating a
 PT skin diseases (e.g. acne, psoriasis and rosacea), tumor diseases,
 PT leukemias, autoimmune diseases, allergies, arthritis, inflammations, or
 PS organ rejections

XX Disclosure; Fig 1; 53pp; English.

CC The present sequence is that of the N-terminal region of human
 CC stearoyl-CoA desaturase (SCD), encoded by the 5' region of the SCD
 CC coding sequence given in ABL57445. The present invention provides
 CC a promoter of the SCD gene and functional moieties, fragments and
 CC variants of it, as well as nucleic acid constructs and vectors that
 CC contain such sequences, and their uses. The promoter may be used
 CC for selective transgenic expression in various tissues such as the
 CC skin for treating a skin disease (e.g. acne, psoriasis and rosacea),
 CC tumours, leukaemia, autoimmune diseases, allergy, arthritis,
 CC inflammation, organ rejection, graft versus host reaction, diseases
 CC of the blood coagulation system, cardiovascular diseases, anaemia,
 CC infections, or damage to the central nervous system.

XX Sequence 16 AA;

Query Match 100.0%; Score 92; DB 23; Length 16;
 Best Local Similarity 100.0%; Pred. No. 5.8e-09;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPAPW 16
 |||||
 DB 1 MPAHLQDDVSFPAPW 16

RESULT 2

AA69378
 ID AAY69378 standard; Protein; 359 AA.

XX AAY69378;

DT 19-JUN-2000 (first entry)

DE Amino acid sequence of human skin stearoyl-CoA desaturase.

XX Mouse; skin; stearoyl-CoA desaturase; SCD; unsaturated fatty acid;
 KW skin disorder; skin cancer; hypertrichosis; hirsutism; acne;
 KW atopic dermatitis; alopecia; gene therapy.

XX Homo sapiens.

OS WO200009754-A2.

PN 24-FEB-2000.

PD 12-AUG-1999; 99WO-US18387.

PF 14-AUG-1998; 98US-0096520.

PR 05-AUG-1999; 99US-0096520.

XX (JOHJ) JOHNSON & JOHNSON CONSUMER CO INC.

XX Stenn K, Prouty SM, Parimoo S, Zhang L;

XX WPI: 2000-224373/19.

DR N-PSDB; AA261576.

XX Novel nucleic acid useful for diagnosing and treating human skin
 PT disorder comprises sequences encoding human stearoyl-CoA desaturase

PS Claim 10; Fig 8; 91pp; English.

XX The present sequence represents a human skin stearoyl-CoA desaturase
 CC (SCD) polypeptide. SCD is responsible for the production of
 CC unsaturated fatty acids, which are required for energy and lipid
 CC metabolism, membrane structure and signal transduction. The SCD
 CC polynucleotides and polypeptides are useful for diagnosing a skin

CC disorder by an abnormal level of SCD expression. The polypeptide is
 CC useful for determining whether an agent increases or decreases the
 CC expression level or activity of human SCD in skin cells. Such compounds
 CC are useful for treating human skin disorders such as skin cancer,
 CC hypertrichosis and hirsutism which is characterized by an excess of
 CC SCD activity. The SCD polypeptides and polynucleotides are also useful
 CC for treating human skin disorders such as acne, atopic dermatitis and
 CC alopecia. The SCD polynucleotide is also useful in gene therapy.

XX Sequence 359 AA;

Query Match 60.9%; Score 56; DB 21; Length 359;
 Best Local Similarity 90.9%; Pred. No. 0.27;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
 |||||
 DB 1 MPAHLQDDIS 11

RESULT 3

ABB44583

ID ABB44583 standard; Protein; 359 AA.

XX ABB44583;

DT 25-JAN-2002 (first entry)

DE Human wound healing related polypeptide SEQ ID NO 40.

XX Human; mouse; vulnery; dermatological; skin disorder; wound healing;
 KW gene therapy.

XX Homo sapiens.

XX CA2325226-A1.

XX 17-MAY-2001.

XX 16-NOV-2000; 2000CA-2325226.

XX 17-NOV-1999; 99DE-1055349.

PR 17-DEC-1999; 99US-0172511.

XX 20-JUN-2000; 2000DE-1030149.

XX (SWIT-) SWITCH BIOTECH AG.

XX Regenbogen J, Wolf E, Goppelt A, Werner S, Halle J;

XX WPI: 2001-433142/47.

XX Use of novel polypeptide or its variant or nucleic acid encoding the
 PT polypeptide for diagnosing and/or preventing and/or treating skin
 PT disorders and/or treatment in wound healing or for identifying active
 PT substances

XX Disclosure; Page 196-198; 265pp; English.

XX The invention relates to the use of a polypeptide (ABB44544-ABB44601,
 CC ABB44606-ABB44623) or its variant or encoding nucleic acid
 CC (ABA91990-ABA81995, ABA82032) with vulnery and/or
 CC dermatological activity for the diagnosis, prevention and treatment of
 CC skin disorders and treatment in wound healing or for the identification
 CC of pharmacologically active substances. The nucleic acids are useful in
 CC gene therapy.

CC Note: The printed sequence listing for this specification was incomplete,
 CC terminating part way through SEQ ID NO 106. The remaining data was
 CC obtained from EPO data for an equivalent patent (EP1114862).

XX Sequence 359 AA;

Query Match 60.9%; Score 56; DB 22; Length 359;
 Best Local Similarity 90.9%; Pred. No. 0.27;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
|||||:|||||
Db 1 MPAHLQDDIS 11

RESULT 4
ABG05270
ID ABG05270 standard; Protein; 285 AA.

XX AC ABG05270;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #5261.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PA 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX N-PSDB; AAS69457.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX PS Claim 20; SEQ ID No 35629; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (I) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 285 AA;

Query Match 53.3%; Score 49; DB 22; Length 285;
Best Local Similarity 61.5%; Pred. No. 3.3;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPW 15
|||||:|||||
Db 223 AHLLETLISFWS 235

RESULT 5
ABG26777

ID ABG26777 standard; Protein; 285 AA.

XX AC ABG26777;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #26768.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PA 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX N-PSDB; AAS90964.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX PS Claim 20; SEQ ID No 57136; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (I) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 285 AA;

Query Match 53.3%; Score 49; DB 22; Length 285;
Best Local Similarity 61.5%; Pred. No. 3.3;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPW 15
|||||:|||||

Db 223 AHLLETLISFPSW 235

RESULT 6

ID ABG06715 standard; Protein; 334 AA.

XX AC ABG06715;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6706.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PS 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS70902.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -

XX PS Claim 20; SEQ ID No 37074; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 334 AA;

Query Match 51.1%; Score 47; DB 22; Length 334;

Best Local Similarity 61.5%; Pred. No. 8.6;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAW 15

|||||:|||||

Db 190 AHLHDLPLISFPSW 202

3

RESULT 7

ABG26818

ID ABG26818 standard; Protein; 334 AA.

XX AC ABG26818;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #26809.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PS 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS91005.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -

XX PS Claim 20; SEQ ID No 57177; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 334 AA;

Query Match 51.1%; Score 47; DB 22; Length 334;

Best Local Similarity 61.5%; Pred. No. 8.6;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAW 15

|||||:|||||

Db 190 AHLHDLPLISFPSW 202

RESULT 8

AAU75109
ID AAU75109 standard; Protein; 1675 AA.
XX
AC AAU75109;
XX
XX
DT 09-APR-2002 (first entry)
XX
DE Clathrin heavy chain.
XX
KW MAPKAP-K3; AP-3 delta; APP-695; amyloid A4; Hsp8; heat shock protein 8;
KW LI130; NY-REN-58; P38 Alpha; P38 Beta; ERK3; KIAA0934; CDK9;
KW cell cycle dependent kinase 9; CLK; C-NAP-1; clathrin heavy chain;
KW amphiphysin; PN9109; KIAA1106; neurodegenerative disorder;
KW non-insulin dependent diabetes mellitus; NIDDM; Alzheimer's disease;
KW inflammatory disease; rheumatoid arthritis; inflammatory bowel disorder;
KW atherosclerosis; cardiac hypertrophy; hypoxic brain injury;
KW yeast two-hybrid; signal transduction pathway; human;
KW mitogen activated protein kinase.
XX
OS Homo sapiens.
XX
FH Key
FT Region 865..1170
FT Location/Qualifiers
FT /note= "this region binds the centrosomal Nek-2
FT associated protein 1 (C-NAP1) (see ABK13313),
FT the bait protein in a yeast two-hybrid assay,
FT producing a complex claimed in claim 1"
XX
XX WO2001198524-A2.
XX
PD 27-DEC-2001.
XX
XX 21-JUN-2001; 2001WO-US19762.
XX
XX 22-JUN-2000; 2000US-213245P.
XX
XX (MYRI-) MYRIAD GENETICS INC.
XX
XX Heichman K, Bartel PL;
PI WPI; 2002-122287/16.
XX
XX New protein complexes comprising protein-protein interactions (e.g.
PT MAPKAP-K3/AP-3 delta or C-NAP-1/Clathrin HC), useful for diagnosing
PT physiological generative disorders or screening drugs for these
PT diseases -
XX
PS Example 10; Page -; 60pp; English.
XX
CC The invention describes an isolated protein complex, comprising two
CC proteins. The protein complex comprises: protein kinase MAPKAP-K3 and
CC AP-3 adaptor complex delta sub-unit; MAPKAP-K3 and amyloid A4
CC precursor protein, APP-695; MAPKAP-K3 and heat shock protein (Hsp) 8;
CC leucine rich LI130 and NY-REN-58; P38 Alpha and P38 Beta; protein kinase
CC ERK3 and KIAA0934 (unknown function); ERK3 and cell cycle dependent
CC kinase (CDK) 9; ERK3 and protein kinase CLK; C-NAP-1 and Clathrin heavy
CC chain; C-NAP-1 and Amphiphysin; C-NAP-1 and novel protein PN9109 or
CC C-NAP-1 and KIAA1106 (unknown function) interactions. The protein
CC complexes are useful for diagnosing physiological generative disorders,
CC drug screening for agents that modulate the interaction of the proteins
CC (thus identify drug targets), and identifying additional proteins in the
CC pathway common to the proteins. These physiological disorders include
CC non-insulin dependent diabetes mellitus (NIDDM), neurodegenerative
CC disorders (e.g. Alzheimer's disease), inflammatory diseases (e.g.
CC rheumatoid arthritis and inflammatory bowel disorder) and other human
CC disease such as atherosclerosis, cardiac hypertrophy and hypoxic brain
CC injury. This sequence represents the clathrin heavy chain protein,
CC residues 865-1170 of which binds to the bait protein centrosomal Nek-2
CC associated protein 1 (C-NAP1) (see ABK13313) in a yeast two-hybrid assay
CC for determining components of signal transduction pathways and forms an
CC interaction claimed in claim 1 of the invention.
CC Note: This sequence does not appear in the specification but has been
CC obtained from a reference given in the invention.

XX SQ Sequence 1675 AA;
Query Match 48.9%; Score 45; DB 23; Length 1675;
Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MPAHLQDDVSFPAP 15
Db 99 MKAHTMTDDVTFWK 113
| | | : | | | : |
RESULT 9
ABP07988
ID ABP07988 standard; Protein; 61 AA.
XX
AC ABP07988;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:15958.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
XX WO2001192523-A2.
XX
PD 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US10836.
XX
XX 30-MAY-2000; 2000US-206132P.
XX
XX 29-AUG-2000; 2000US-228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach MD;
PI WPI; 2002-106308/14.
XX
XX N-PSDB; ABN23740.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders -
XX
XX Disclosure; SEQ ID 15958; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,

D_b 171 IPGHIPGKVKEPVGFPMW 188

Query Match 47.3%; Score 43.5; DB 23; Length 632;
 Best Local Similarity 42.9%; Pred. No. 70;
 Matches 9; Conservative 4; Mismatches 1; Indels 7; Gaps 1;

QY 1 MPAHL-----LQDDVSEPA 14
 || || |:::|||||
 Db 430 MPGLVNASYDPLKNDITFPA 450

RESULT 12
 ABP26786
 ID ABP26786 standard; Protein; 437 AA.
 XX AC ABP26786;
 XX DT
 XX DT 02-JUL-2002 (first entry)
 XX DE
 XX DE Streptococcus polypeptide SEQ ID NO 2748.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX OS Streptococcus pyogenes.
 XX PN WO200234771-A2.
 XX PD
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB04789.
 XX PR 27-OCT-2000; 2000GB-0026333.
 XX PR 24-NOV-2000; 2000GB-0028727.
 XX PR 07-MAR-2001; 2001GB-0005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX WPI; 2002-352536/38.
 XX N-PSDB; ABN67417.
 XX DR
 XX DR New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 XX Claim 1; Page 3426; 4525pp; English.
 XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX SQ Sequence 437 AA;

Query Match 46.7%; Score 43; DB 23; Length 437;
 Best Local Similarity 54.5%; Pred. No. 56;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LQDDVSEPAW 15
 ::::|||||
 Db 117 IIQDDISESPW 127

RESULT 13
 ABP33455
 ID ABP33455 standard; Protein; 96 AA.
 XX AC ABP33455;
 XX DT
 XX DT 08-JUL-2002 (first entry)
 XX DE
 XX DE Human ORF2429 protein, SEQ ID NO:4858.
 XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 XX OS Homo sapiens.
 XX PN WO200190366-A2.
 XX XX 29-NOV-2001.
 XX PF 24-MAY-2001; 2001WO-US17076.
 XX PR 24-MAY-2000; 2000US-206690P.
 XX PA (CJRA-) CURAGEN CORP.
 XX PI Leach MD, Shimkets RA;
 XX WPI; 2002-106200/14.
 XX N-PSDB; ABN77482.
 XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation -
 XX Claim 10; Page 1468; 2508pp; English.
 XX CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,

CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.

XX
 XX Sequence 96 AA;

Query Match 45.7%; Score 42; DB 23; Length 96;

Best Local Similarity 50.0%; Pred. No. 15; Indels 7; Gaps 0;
 Matches 8; Conservative 1; Mismatches 7;

QY 1 MPAHLQDDVSEFPWP 16
 I III I: III
 Db 60 METHLLVSRVTMAAWP 75

RESULT 14

AAAY70265
 ID AAY70265 standard; Protein; 184 AA.

XX
 AC AAY70265;

XX
 DT 06-JUN-2000 (first entry)

XX
 DE Canine angiogenesis inhibitor, endostatin.

XX
 DE Canine; immunoglobulin Fc fragment; endostatin; immunofusin;
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
 KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
 KW vasotropic; vulnerable; treatment; antiarteriosclerosis; tumour;
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KW wound granulation; keloid scar; gene therapy.

XX
 OS Canis familiaris.

XX
 PN W0200011033-A2.

XX
 PD 02-MAR-2000.

XX
 PF 25-AUG-1999; 99WO-US19329.

XX
 PR 25-AUG-1998; 98US-0097883.

XX
 PA (LEXI-) LEXINGEN PHARM CORP.

XX
 PI Lo K, Li Y, Gillies SD;

XX
 DR WPI; 2000-237616/20.

XX
 DR N-PSDB; AAZ51309.

XX
 PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin
 FC region, useful for treating conditions mediated by angiogenesis,
 PT such as rheumatoid arthritis, tumors and macular degeneration -

XX
 PS Example 8; Pages 59-60; 68pp; English.

XX
 CC The patent discloses a DNA molecule encoding a fusion protein comprising
 a signal sequence, an immunoglobulin Fc region, and an angiogenesis
 CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment
 CC having angiostatin activity, a collagen XVIII fragment having endostatin

CC activity, or combinations of them. The fusion protein (immunofusin) is
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated
 CC by angiogenesis. Conditions that may be treated include solid tumours,
 CC blood born tumours, tumour metastasis, benign tumours including
 CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, rubecosis and Osler-Webber syndrome; myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
 CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
 CC in gene therapy. The present sequence is a canine
 CC endostatin used in the construction of immunofusin containing canine
 CC immunoglobulin Fc fragment.

XX
 SQ Sequence 184 AA;

Query Match 45.7%; Score 42; DB 21; Length 184;

Best Local Similarity 46.7%; Pred. No. 32;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSEFPWP 15

Db 59 VPVNLDEVLFPSW 83

RESULT 15

AAO17430
 ID AAO17430 standard; Protein; 184 AA.

XX
 AC AAO17430;

XX
 DT 19-JUL-2002 (first entry)

XX
 DE Canine endostatin.

XX
 DE Dog; pro-endostatin; endostatin; angiogenesis; cancer; gene therapy;
 KW psoriasis; rheumatoid arthritis; retinopathy; macular degeneration;
 KW corneal graft rejection; neovascular glaucoma; retrolental fibroplasia;
 KW rubecosis; Osler-Webber syndrome; myocardial angiogenesis;
 KW plaque neovascularisation; telangiectasia; haemophilic joints;
 KW angiofibroma; wound granulation; coronary collateral;
 KW cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; diabetic neovascularisation; fracture;
 KW cytostatic; antiarthritic; antipsoriatic; antidiabetic; ophthalmological;
 KW gynaecological.

XX
 OS Canis familiaris.

XX
 PN EP1191036-A2.

XX
 PD 27-MAR-2002.

XX
 PF 24-AUG-2001; 2001EP-0307224.

XX
 PR 25-AUG-2000; 2000US-227924P.

XX
 PA (PFIZ) PFIZER PROD INC.

XX
 PI Sheppard MG, Tong X;

XX
 DR WPI; 2002-354068/39.

XX
 DR N-PSDB; AAL46063.

XX
 PT An isolated nucleic acid molecule for the treatment of
 angiogenesis-related disorder, such as cancers or diabetic retinopathy,
 PT encodes an endostatin protein -

XX
 PS Claim 14; Fig 5; 56pp; English.

XX
 CC The present invention provides the protein and coding sequences of canine

3333

2

5

1

22

2

Time : 57 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:23:29 ; Search time 14 Seconds
(without alignments)
33.626 Million cell updates/sec

Title: US-10-016-725-16
Perfect score: 92
Sequence: 1 MPAHLLQDDVSFPAPW 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued Patents_AA.*
- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40	43.5	68	4	US-08-441-507-9
2	40	43.5	68	4	US-07-969-875A-9
3	40	43.5	86	4	US-08-441-507-8
4	40	43.5	86	4	US-07-969-875A-8
5	40	43.5	197	4	US-08-441-507-5
6	40	43.5	197	4	US-07-969-875A-5
7	40	43.5	200	4	US-08-441-507-4
8	40	43.5	200	4	US-07-969-875A-4
9	40	43.5	246	4	US-08-441-507-21
10	40	43.5	246	4	US-07-969-875A-21
11	40	43.5	272	4	US-08-441-507-15
12	40	43.5	272	4	US-07-969-875A-15
13	39	42.4	356	1	US-08-552-142A-15
14	39	42.4	356	1	US-08-910-973-15
15	39	42.4	356	3	US-09-234-332-7
16	39	42.4	356	3	US-09-234-332-8
17	39	42.4	356	4	US-09-499-227-15
18	39	42.4	357	1	US-08-552-142A-2
19	39	42.4	357	4	US-08-910-973-2
20	39	42.4	357	4	US-09-499-227-2
21	39	42.4	357	5	PCI-US95-05741-2
22	39	42.4	509	4	US-08-857-076-46
23	39	42.4	510	4	US-08-857-076-45
24	39	42.4	635	4	US-08-857-076-101
25	38	41.3	38	1	US-08-176-500-49
26	38	41.3	38	1	US-08-471-052A-49
27	38	41.3	38	1	US-08-189-331-49

28	38	41.3	38	2	US-08-471-939-49	Sequence 49, Appl
29	38	41.3	38	2	US-08-471-800-49	Sequence 49, Appl
30	38	41.3	38	2	US-08-471-068-49	Sequence 49, Appl
31	38	41.3	178	4	US-09-315-689-5	Sequence 5, Appl
32	38	41.3	182	4	US-09-561-500-14	Sequence 14, Appl
33	38	41.3	182	4	US-09-561-108-14	Sequence 14, Appl
34	38	41.3	182	4	US-09-315-689-3	Sequence 3, Appl
35	38	41.3	182	4	US-09-561-526-14	Sequence 14, Appl
36	38	41.3	183	4	US-09-206-059-2	Sequence 2, Appl
37	38	41.3	662	4	US-09-061-768A-25	Sequence 25, Appl
38	37	40.2	47	4	US-09-149-476-359	Sequence 359, App
39	37	40.2	216	4	US-09-077-955-31	Sequence 31, Appl
40	37	40.2	256	4	US-09-077-955-30	Sequence 30, Appl
41	37	40.2	294	4	US-09-077-955-29	Sequence 29, Appl
42	37	40.2	338	4	US-09-077-955-28	Sequence 28, Appl
43	37	40.2	358	2	US-08-558-823-19	Sequence 19, Appl
44	37	40.2	368	2	US-08-869-137-2	Sequence 2, Appl
45	37	40.2	390	4	US-09-077-955-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-441-507-9
; Sequence 9, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
; TITLE OF INVENTION: Dactylon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-441-507-9
Query Match 43.5%; Score 40; DB 4; Length 68;
Best Local Similarity 61.5%; Pred. No. 4.9;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLLODDVSPAW 15
Db 42 AHLVQDDVIPANW 54

RESULT 2
US-07-969-875A-9
; Sequence 9, Application US/07969875A
; Patent No. 6441157
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species
; TITLE OF INVENTION: Cynodon Dactylon
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vanstone, Darlene A.;
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-07-969-875A-9

Query Match 43.5%; Score 40; DB 4; Length 68;
Best Local Similarity 61.5%; Pred. No. 4.9;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLLODDVSPAW 15
Db 42 AHLVQDDVIPANW 54

RESULT 3
US-08-441-507-8
; Sequence 8, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species
; TITLE OF INVENTION: Dactylon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts

COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,507
FILING DATE: 15-May-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/969,875
FILING DATE: 30-October-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-049DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 86 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-441-507-8

Query Match 43.5%; Score 40; DB 4; Length 86;
Best Local Similarity 61.5%; Pred. No. 6.4;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLLODDVSPAW 15
Db 50 AHLVQDDVIPANW 72

RESULT 4
US-07-969-875A-8
; Sequence 8, Application US/07969875A
; Patent No. 6441157
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species
; TITLE OF INVENTION: Cynodon Dactylon
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vanstone, Darlene A.;
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010

; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 86 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; US-07-969-875A-8

Query Match 43.5%; Score 40; DB 4; Length 86;
 Best Local Similarity 61.5%; Pred. No. 6.4;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAP 15
 DB 60 AHLVQDDVIPANW 72

RESULT 5

; US-08-441-507-5
 ; Sequence 5, Application US/08441507
 ; Patent No. 6214358
 ; GENERAL INFORMATION:
 ; APPLICANT: Singh, Mohan Bir;
 ; APPLICANT: Smith, Penelope; and
 ; APPLICANT: Knox, Robert Bruce
 ; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
 ; TITLE OF INVENTION: Dactylon
 ; NUMBER OF SEQUENCES: 52
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD LLP
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/441,507
 ; FILING DATE: 15-May-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/969,875
 ; FILING DATE: 30-October-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandragouras, Amy E.
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: IMI-049DV
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 227-7400
 ; TELEFAX: (617) 742-4214

; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 197 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; US-08-441-507-5

Query Match 43.5%; Score 40; DB 4; Length 197;
 Best Local Similarity 61.5%; Pred. No. 17;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAP 15
 DB 173 AHLVQDDVIPANW 185

RESULT 6

; US-07-969-875A-5
 ; Sequence 5, Application US/07969875A
 ; Patent No. 6441157
 ; GENERAL INFORMATION:
 ; APPLICANT: Singh, Mohan Bir;
 ; APPLICANT: Smith, Penelope; and
 ; APPLICANT: Knox, Robert Bruce
 ; TITLE OF INVENTION: Protein Allergens of the Species
 ; TITLE OF INVENTION: Cynodon dactylon
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Immunologic Pharmaceutical Corporation
 ; STREET: 610 Lincoln Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02154

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/969,875A
 ; FILING DATE: 30-October-1992
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vanstone, Darlene A..
 ; REGISTRATION NUMBER: 35,279
 ; REFERENCE/DOCKET NUMBER: 041.1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 466-6000
 ; TELEFAX: (617) 466-6010
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 197 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; US-07-969-875A-5

Query Match 43.5%; Score 40; DB 4; Length 197;
 Best Local Similarity 61.5%; Pred. No. 17;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSPFAP 15
 DB 173 AHLVQDDVIPANW 185

RESULT 7

; US-08-441-507-4
 ; Sequence 4, Application US/08441507
 ; Patent No. 6214358
 ; GENERAL INFORMATION:
 ; APPLICANT: Singh, Mohan Bir;
 ; APPLICANT: Smith, Penelope; and
 ; APPLICANT: Knox, Robert Bruce
 ; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
 ; TITLE OF INVENTION: Dactylon
 ; NUMBER OF SEQUENCES: 52
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD LLP
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-441-507-4

```

```

Query Match 43.5%; Score 40; DB 4; Length 200;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 AHLQDDVSFPANW 15
   |||:||||
Db 174 AHLVQDDVIPANW 186

```

RESULT 8

```

US-07-969-875A-4
; Sequence 4, Application US/07969875A
; Patent No. 6441157
; GENERAL INFORMATION:

```

```

; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species
; TITLE OF INVENTION: Cynodon Dactylon
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vanstone, Darlene A.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-07-969-875A-4

```

Query Match 43.5%; Score 40; DB 4; Length 200;
Best Local Similarity 61.5%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 AHLQDDVSFPANW 15
   |||:||||
Db 174 AHLVQDDVIPANW 186

```

RESULT 9

```

US-08-441-507-21
; Sequence 21, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:

```

```

; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
; TITLE OF INVENTION: Dactylon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-441-507-21

```

```

Query Match 43.5%; Score 40; DB 4; Length 246;
Best Local Similarity 61.5%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 AHLQDDVSFPANW 15
   |||:||||
Db 220 AHLVQDDVIPANW 232

```

RESULT 10

```

US-07-969-875A-21
; Sequence 21, Application US/07969875A
; Patent No. 6441157
; GENERAL INFORMATION:

```

```

; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce

```

```

; TITLE OF INVENTION: Protein Allergens of the Species
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vanstone, Darlene A.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-969-875A-21

Query Match 43.5%; Score 40; DB 4; Length 246;
Best Local Similarity 61.5%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPW 15
Db 220 AHLVDDVIPANW 232
|||||
|

RESULT 11
US-08-441-507-15
; Sequence 15, Application US/08441507
; Patent No. 6214358
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species Cynodon
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,507
; FILING DATE: 15-May-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,875
; FILING DATE: 30-October-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 35,207
; REFERENCE/DOCKET NUMBER: IMI-049DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5-8
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15-16
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 42
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 71-72
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; US-08-441-507-15

Query Match 43.5%; Score 40; DB 4; Length 272;
Best Local Similarity 61.5%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPW 15
Db 246 AHLVDDVIPANW 258
|||||
|

RESULT 12
US-07-969-875A-15
; Sequence 15, Application US/07969875A
; Patent No. 6441157
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir;
; APPLICANT: Smith, Penelope; and
; APPLICANT: Knox, Robert Bruce
; TITLE OF INVENTION: Protein Allergens of the Species
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation
; STREET: 610 Lincoln Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/969,875A
; FILING DATE: 30-October-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Vanstone, Darlene A.
; REGISTRATION NUMBER: 35,279
; REFERENCE/DOCKET NUMBER: 041.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5-8
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15-16
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 42
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 71-72
; OTHER INFORMATION: /note= "Xaa is an unknown amino acid"
; US-07-969-875A-15

```

```

Query Match 43.5%; Score 40; DB 4; Length 272;
Best Local Similarity 61.5%; Pred. NO. 25;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 3 AHLQDDVSPFPAW 15
    |||:||||
Db 246 AHLVQDDVIPANW 258

```

```

RESULT 13
US-08-552-142A-15
; Sequence 15, Application US/08552142A
; Patent No. 5695995
; GENERAL INFORMATION:
; APPLICANT: Weintraub, Harold M.
; APPLICANT: Lee, Jacqueline E.
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Hollenberg, Stanley M.
; TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/552,142A
; FILING DATE: 02-NOV-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: FHCR-1-8933
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-225-0709
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-552-142A-15

```

```

Query Match 42.4%; Score 39; DB 1; Length 356;
Best Local Similarity 50.0%; Pred. NO. 52;
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 MPAHLQDDVSPFPAW 16
    |||:||||
Db 203 MPHLPTASAFVHP 218

```

```

RESULT 14
US-08-910-973-15
; Sequence 15, Application US/08910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectod
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 15:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-910-973-15

Query Match 42.4%; Score 39; DB 1; Length 356;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 16
|||
Db 203 MPPLPTASFPVHP 218

RESULT 15

US-09-234-332-7
; Sequence 7, Application US/09234332A
; Patent No. 6087168
; GENERAL INFORMATION:
; APPLICANT: Cedars-Sinai Medical Center
; APPLICANT: Michel F. Levesque, M.D.
; APPLICANT: Thomas Neuman, Ph.D.
; TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
; FILE REFERENCE: P07 41494
; CURRENT APPLICATION NUMBER: US/09/234.332A
; CURRENT FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (0)...(0)
; OTHER INFORMATION: Neuro D1 protein; Genbank Accession D82347
US-09-234-332-7

Query Match 42.4%; Score 39; DB 3; Length 356;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 16
|||
Db 203 MPPLPTASFPVHP 218

Search completed: May 14, 2003, 15:31:09
Job time : 15 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:20:36 ; Search time 17 seconds
(without alignments)
86.612 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92

Sequence: 1 MPAHLQDDVSFPAMP 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	92	100.0	16	12	US-10-016-725-16
2	56	60.9	359	9	US-10-060-036-179
3	56	60.9	359	12	US-10-029-654-12
4	44	47.8	71	9	US-10-001-857-179
5	43.5	47.3	632	9	US-09-971-536-42
6	42	45.7	184	9	US-10-131-241-49
7	42	45.7	255	9	US-10-224-446-28
8	41	44.6	737	9	US-10-228-931-4
9	41	44.6	737	10	US-09-771-161A-195
10	40	43.5	81	9	US-09-738-626-6569
11	40	43.5	128	9	US-09-764-891-2789
12	40	43.5	255	9	US-10-224-446-30
13	39.5	42.9	72	10	US-09-864-761-42012
14	39	42.4	74	9	US-09-910-009A-56
15	39	42.4	74	9	US-09-910-009A-74
16	39	42.4	74	9	US-09-910-009A-138
17	39	42.4	74	9	US-09-910-009A-144
18	39	42.4	74	9	US-09-910-009A-147
19	39	42.4	75	9	US-09-910-009A-201

20	39	42.4	75	10	US-09-764-877-1086	Sequence 1086, Ap
21	39	42.4	76	9	US-09-910-009A-265	Sequence 265, App
22	39	42.4	78	9	US-09-910-009A-14	Sequence 14, Appl
23	39	42.4	78	9	US-09-910-009A-178	Sequence 178, App
24	39	42.4	102	9	US-09-925-299-877	Sequence 877, App
25	39	42.4	102	10	US-09-925-299-877	Sequence 877, App
26	39	42.4	155	9	US-10-121-988-181	Sequence 181, App
27	39	42.4	396	10	US-09-864-761-43232	Sequence 43232, A
28	39	42.4	509	10	US-09-205-658-46	Sequence 46, Appl
29	39	42.4	509	10	US-09-844-353A-46	Sequence 46, Appl
30	39	42.4	510	10	US-09-205-658-45	Sequence 45, Appl
31	39	42.4	510	10	US-09-844-353A-45	Sequence 45, Appl
32	39	42.4	635	10	US-09-844-353A-101	Sequence 101, App
33	39	42.4	704	9	US-10-121-988-94	Sequence 94, Appl
34	39	42.4	858	9	US-10-121-988-92	Sequence 92, Appl
35	39	42.4	1293	10	US-09-815-242-10079	Sequence 10079, A
36	39	42.4	1294	10	US-09-815-242-13724	Sequence 13724, A
37	38	41.3	63	10	US-09-822-540A-1	Sequence 1, Appl
38	38	41.3	101	10	US-09-864-761-45805	Sequence 45805, A
39	38	41.3	178	9	US-10-131-241-60	Sequence 60, Appl
40	38	41.3	179	9	US-10-131-241-57	Sequence 57, Appl
41	38	41.3	180	9	US-10-131-241-47	Sequence 47, Appl
42	38	41.3	180	9	US-10-131-241-56	Sequence 56, Appl
43	38	41.3	181	9	US-10-131-241-55	Sequence 55, Appl
44	38	41.3	182	9	US-10-131-241-54	Sequence 54, Appl
45	38	41.3	182	10	US-09-998-831-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-016-725-16
; Sequence 16, Application US/10016725
; Patent No. US20020151018A1
; GENERAL INFORMATION:
; APPLICANT: Prouty, Stephen
; APPLICANT: Zhang, Lin
; APPLICANT: Stenn, Kurt
; TITLE OF INVENTION: Stearyl-CoA Desaturase Promoter
; FILE REFERENCE: J4J2065
; CURRENT APPLICATION NUMBER: US/10/016,725
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-725-16

Query Match 100.0%; Score 92; DB 12; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.8e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPAMP 16
Db 1 MPAHLQDDVSFPAMP 16

RESULT 2
US-10-050-036-179
; Sequence 179, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; AND DIAGNOSIS OF PANCREATIC CANCER

; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060.036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-179

Query Match 60.9%; Score 56; DB 9; Length 359;
Best Local Similarity 90.9%; Pred. No. 0.2;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
||| ||||| :||
Db 1 MPAHLQDDIS 11

RESULT 3
US-10-029-654-12
; Sequence 12, Application US/10029654
; Patent No. US20020150958A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim Pharma KG
; TITLE OF INVENTION: Methods for identifying substances for treating
; TITLE OF INVENTION: inflammatory conditions
; FILE REFERENCE: 1/1178
; CURRENT APPLICATION NUMBER: US/10/029,654
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/257,878
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-654-12

Query Match 60.9%; Score 56; DB 12; Length 359;
Best Local Similarity 90.9%; Pred. No. 0.2; Indels 0; Gaps 0;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
||| ||||| :||
Db 1 MPAHLQDDIS 11

RESULT 4
US-10-001-857-179
; Sequence 179, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 179
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-857-179

Query Match 47.8%; Score 44; DB 9; Length 71;
Best Local Similarity 61.5%; Pred. No. 3;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 4 HLLQDDVSFPAPW 16
|| | || | :||
Db 38 HLSGDRVSTPSPW 50

RESULT 5
US-09-971-536-42
; Sequence 42, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
; TITLE OF INVENTION: Using Them
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-42

Query Match 47.3%; Score 43.5; DB 9; Length 632;
Best Local Similarity 42.9%; Pred. No. 44; Indels 7; Gaps 1;
Matches 9; Conservative 4; Mismatches 1; Indels 7; Gaps 1;

QY 1 MPAHL-----LQDDVSFPA 14
|| | | :|| :||
Db 430 MPGLVNASYDPLKNDITEPA 450

RESULT 6
US-10-131-241-49
; Sequence 49, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holladay, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolif
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 49
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Canine sp.
US-10-131-241-49

Query Match 45.7%; Score 42; DB 9; Length 184;
Best Local Similarity 46.7%; Pred. No. 19;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 MPAHLQDDVSFPAP 15
:| | | | | | | | | |
Db 69 VPVNLRLDEVLPFSW 83

RESULT 7
US-10-224-446-28
; Sequence 28, Application US/10224446
; Publication No. US20030024010A1
; GENERAL INFORMATION:
; APPLICANT: CAHOON, EDGAR B.
; APPLICANT: CARLSON, THOMAS J.
; APPLICANT: HITZ, WILLIAM D.
; APPLICANT: RIPP, KEVIN G.
; TITLE OF INVENTION: GENES FOR PLANT FATTY ACID MODIFYING ENZYMES
; TITLE OF INVENTION: ASSOCIATED
; TITLE OF INVENTION: WITH CONJUGATED DOUBLE BOND FORMATION
; FILE REFERENCE: BB-1249
; CURRENT APPLICATION NUMBER: US/10/224,446
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US/09/375,299
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/097,186
; PRIOR FILING DATE: August 20, 1998
; PRIOR APPLICATION NUMBER: 60/142,756
; PRIOR FILING DATE: July 8, 1999
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 28
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Aleurites fordii
US-10-224-446-28

Query Match 45.7%; Score 42; DB 9; Length 255;
Best Local Similarity 53.8%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSFPAP 16
:| | | | | | | | | |
Db 38 HLLPQLSYVAMP 50

RESULT 8
US-10-228-931-4
; Sequence 4, Application US/10228931
; Publication No. US20030031259A1
; GENERAL INFORMATION:
; APPLICANT: Verma, Ajit K
; APPLICANT: Reddig, Peter J
; APPLICANT: Jansen, Aaron P
; TITLE OF INVENTION: Animal Model System for Squamous Cell Carcinoma
; FILE REFERENCE: 960296.97613
; CURRENT APPLICATION NUMBER: US/10/228,931
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US/09/772,647
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T7 tag and
; OTHER INFORMATION: mouse protein kinase C epsilon coding sequence
US-10-228-931-4

Query Match 44.6%; Score 41; DB 9; Length 737;
Best Local Similarity 54.5%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 5 LLQDDVSFPAP 15
:| | | | | | | | | |
Db 619 ILHDDVLYPVW 629

RESULT 9
US-09-771-161A-195
; Sequence 195, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 803620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 195
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-195

Query Match 44.6%; Score 41; DB 10; Length 737;
Best Local Similarity 54.5%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPAP 15
:| | | | | | | | | |
Db 619 ILHDDVLYPVW 629

RESULT 10
US-09-738-626-6569
; Sequence 6569, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6569
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6569

Query Match 43.5%; Score 40; DB 9; Length 81;
Best Local Similarity 35.7%; Pred. No. 16;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSPFA 14
Db 7 LFNHLHRTDITPS 20

RESULT 11
US-09-764-891-2789
; Sequence 2789, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2789
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-2789

Query Match 43.5%; Score 40; DB 9; Length 128;
Best Local Similarity 54.5%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 LODDVSPFPAW 16
Db 12 LSPETGFPAPW 22

RESULT 12
US-10-224-446-30
; Sequence 30, Application US/10224446
; Publication No. US20030024010A1
; GENERAL INFORMATION:
; APPLICANT: CAHOON, EDGAR B.
; APPLICANT: CARLSON, THOMAS J.
; APPLICANT: HITZ, WILLIAM D.
; APPLICANT: RIPP, KEVIN G.
; TITLE OF INVENTION: GENES FOR PLANT PATTY ACID MODIFYING ENZYMES
; TITLE OF INVENTION: ASSOCIATED
; TITLE OF INVENTION: WITH CONJUGATED DOUBLE BOND FORMATION
; FILE REFERENCE: BB-1249
; CURRENT APPLICATION NUMBER: US/10/224,446
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US/09/375,299
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/097,186
; PRIOR FILING DATE: August 20, 1998
; PRIOR APPLICATION NUMBER: 60/142,756
; PRIOR FILING DATE: July 8, 1999
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 30
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Aleurites fordii
US-10-224-446-30

Query Match 43.5%; Score 40; DB 9; Length 255;
Best Local Similarity 46.2%; Pred. No. 58;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSPFPAW 16
Db 38 HLLPSPITYIAMP 50

RESULT 13
US-09-864-761-42012
; Sequence 42012, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42012
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL137190.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7

```

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.3
; OTHER INFORMATION: SWISSPROT HIT: O84337, EVALUE 4.70e+00
; OTHER INFORMATION: EST_HUMAN HIT: R78762.1, EVALUE 2.40e+00
US-09-864-761-42012

```

```

Query Match          42.9%; Score 39.5; DB 10; Length 72;
Best Local Similarity 44.4%; Pred. NO. 17;
Matches 8; Conservative 3; Mismatches 4; Indels 4; Gaps 1;

```

```

QY 2 PAHL--LQDDVSFPWP 16
    ||| ||||: ||
Db 38 PSHLKKPLQDEIPRGWP 55

```

RESULT 14

```

US-09-910-009A-56
; Sequence 56, Application US/09910009A
; Publication No. US20030050234A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; APPLICANT: Olivera, Baldomero M.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Garrett, James E.
; APPLICANT: Watkins, Maren
; APPLICANT: Cruz, Lourdes J.
; APPLICANT: Shon, Ki-Joon
; APPLICANT: Jacobsen, Richard
; APPLICANT: Jones, Robert M.
; APPLICANT: Cartier, G. Edward
; APPLICANT: Shen, Greg S.
; APPLICANT: Wagstaff, John D.
; TITLE OF INVENTION: Mu-Conopeptides
; FILE REFERENCE: 2314-242
; CURRENT APPLICATION NUMBER: US/09/910,009A
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 60/219,619
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/245,157
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/264,319
; PRIOR FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 60/277,270
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 520
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Conus characteristicus
US-09-910-009A-56

```

```

Query Match          42.4%; Score 39; DB 9; Length 74;
Best Local Similarity 46.7%; Pred. NO. 21;
Matches 7; Conservative 3; Mismatches 5; Indels 5; Gaps 0;

```

```

QY 2 PAHLQDDVSFPWP 16
    || :|||: ||
Db 32 PADRMQDDISSEQYP 46

```

RESULT 15

```

US-09-910-009A-74
; Sequence 74, Application US/09910009A
; Publication No. US20030050234A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Research Foundation
; APPLICANT: Cognetix, Inc.
; APPLICANT: Olivera, Baldomero M.
; APPLICANT: McIntosh, J. Michael
; APPLICANT: Garrett, James E.
; APPLICANT: Watkins, Maren

```

```

; APPLICANT: Cruz, Lourdes J.
; APPLICANT: Shon, Ki-Joon
; APPLICANT: Jacobsen, Richard
; APPLICANT: Jones, Robert M.
; APPLICANT: Cartier, G. Edward
; APPLICANT: Shen, Greg S.
; APPLICANT: Wagstaff, John D.
; TITLE OF INVENTION: Mu-Conopeptides
; FILE REFERENCE: 2314-242
; CURRENT APPLICATION NUMBER: US/09/910,009A
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: US 60/219,619
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/245,157
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/264,319
; PRIOR FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US 60/277,270
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 520
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Conus circumciscus
US-09-910-009A-74

```

```

Query Match          42.4%; Score 39; DB 9; Length 74;
Best Local Similarity 46.7%; Pred. NO. 21;
Matches 7; Conservative 3; Mismatches 5; Indels 5; Gaps 0;

```

```

QY 2 PAHLQDDVSFPWP 16
    || :|||: ||
Db 32 PADRMQDDISSEQYP 46

```

```

Search completed: May 14, 2003, 15:30:19
Job time : 26 secs

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:30:26 ; Search time 44 seconds
(without alignments)
34.958 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92
Sequence: 1 MPAHLQDDVSPAWP 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	48.9	1675	1 LRTH	clathrin heavy cha
2	44	47.8	1687	2 T43144	vitellogenin II pr
3	43.5	47.3	518	2 T01318	pectinesterase hom
4	43	46.7	136	2 S72870	hypothetical prote
5	43	46.7	414	2 E84653	probable poly(A) p
6	42	45.7	663	1 A35087	arachidonate 12-li
7	42	45.7	1280	2 T51500	hypothetical prote
8	41.5	45.1	245	2 S39481	glutamate-ammonia
9	41.5	45.1	357	2 S39477	glutamate-ammonia
10	41.5	45.1	627	2 H82573	exonuclease ABC s
11	41	44.6	160	2 A84104	hypothetical prote
12	41	44.6	227	2 A70885	hypothetical prote
13	41	44.6	338	2 T21161	hypothetical prote
14	41	44.6	387	2 F82746	DNA processing cha
15	41	44.6	547	2 T22856	hypothetical prote
16	41	44.6	642	2 AH2715	hypothetical prote
17	41	44.6	660	2 AD0661	exonuclease ABC c
18	41	44.6	681	2 C97497	invasin-like prote
19	41	44.6	685	2 A83408	exonuclease ABC c
20	41	44.6	707	1 A53530	exonuclease ABC c
21	41	44.6	736	1 KIRBCE	protein kinase C (
22	41	44.6	737	1 S28942	protein kinase C (
23	41	44.6	737	1 KIRTCE	protein kinase C (
24	41	44.6	737	1 KIMSCC	protein kinase C (
25	41	44.6	847	2 G85666	hypothetical prote
26	41	44.6	847	2 AH0636	periplasmic glucan
27	41	44.6	847	2 F84847	glucan biosynthesi
28	41	44.6	857	2 C90807	membrane glycosyl
29	41	44.6	1146	2 F84487	probable ABC trans

30 41 44.6 2535 2 T02646
31 40.5 44.0 356 2 T03253
32 40.5 44.0 844 2 S77383
33 40 43.5 175 1 G69518
34 40 43.5 243 1 JJAG32
35 40 43.5 243 2 S03122
36 40 43.5 243 2 AB3248
37 40 43.5 320 2 E82520
38 40 43.5 377 2 T04585
39 40 43.5 522 2 G96526
40 40 43.5 1042 2 S76045
41 40 43.5 1089 2 S53978
42 40 43.5 1102 2 H84545
43 39.5 42.9 374 2 F87596
44 39 42.4 108 2 B84229
45 39 42.4 130 2 T08329

ALIGNMENTS

RESULT 1

LRTH

clathrin heavy chain - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 22-Jun-1999

C:Accession: A39941

R:Kirchhausen, T.; Harrison, S.C.; Ping Chow, E.; Mattaliano, R.J.; Ramachandran, K.; Proc. Natl. Acad. Sci. U.S.A. 84, 8805-8809, 1987

A:Title: Clathrin heavy chain: molecular cloning and complete primary structure.

A:Reference number: A39941; MUID:88097376; PMID:3480512

A:Accession: A39941

A:Molecule type: mRNA

A:Residues: 1-1675 <KIR>

A:Cross-references: GB:J03583; NID:g203301; PIDN:AAA0874.1; PID:g203302

C:Comment: Clathrin, the major protein component of coated pits and vesicles, is a t r its carboxyl end. The heavy chains are also held together by noncovalent interacti

C:Comment: The amino end of the mature protein is blocked.

C:Superfamily: clathrin heavy chain

C:Keywords: coated pits

F:1-479/Domain: amino-terminal <TER>

F:480-523/Region: link

F:524-634/Domain: distal <DIS>

F:635-638/Region: joint #status predicted

F:639-1675/Domain: proximal <PRX>

Query Match

Best Local Similarity 48.9%; Score 45; DB 1; Length 1675;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSPAW 15

Db 99 MKAHMTDDVTFMKW 113

RESULT 2

T43144

vitellogenin II precursor - mummichog

C:Species: Fundulus heteroclitus (mummichog)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000

C:Accession: T43144

R:Laflaur, G.J.; Hoch, K.L.; Denslow, N.; Byrne, B.M.; Wallace, R.A.

submitted to the EMBL Data Library, September 1996

A:Description: Derivation of oocyte and egg proteins from parental vitellogenins in 1 n.

A:Reference number: Z22316

A:Accession: T43144

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1687 <LAF>

A:Cross-references: EMBL:U70826; NID:g1621358; PID:g1621359; PIDN:AAB17152.1

A:Experimental source: estrogen-induced liver

C:Superfamily: vitellogenin

C:Keywords: egg yolk; phosphoprotein
F:1-15/Domain: signal sequence status predicted <Sig>
F:16-1687/Product: vitellogenin II #status predicted <MAT>

Query Match 47.8%; Score 44; DB 2; Length 1687;
Best Local Similarity 61.5%; Pred. No. 64;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPW 14
|||||: 1-1 |
DB 831 PAHLKSDISMKR 843

RESULT 3
T01318
pectinesterase homolog T14P8.1 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 24-Sep-1999
C:Accession: T01318
R:Kalicki, J.; Elliott, G.; Cloud, J.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of A. thaliana T14P8.
A:Reference number: Z14290
A:Accession: T01318
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-518 <KAL>
A:Cross-references: EMBL:AF069298; NID:g3193282; PIDN:AAC19272.1; PID:g3193288
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 235/2; 289/1
A:Note: T14P8.1
C:Superfamily: pectinesterase

Query Match 47.3%; Score 43.5; DB 2; Length 518;
Best Local Similarity 38.9%; Pred. No. 20;
Matches 7; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 MPAHL---LQDDVSFPW 15
:| |: :||| |
DB 171 IPGHIPGKVKEDVGPPW 188

RESULT 4
S72870
hypothetical protein B2126_Fl_18 - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S72870
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid B2126.
A:Reference number: S72585
A:Accession: S72870
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-136 <SMI>
A:Cross-references: EMBL:U00017; NID:g466994; PIDN:AAA17210.1; PID:g467025
C:Superfamily: Mycobacterium leprae hypothetical protein B2126_Fl_18

Query Match 46.7%; Score 43; DB 2; Length 136;
Best Local Similarity 70.0%; Pred. No. 5.1;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAHLQDDVS 11
|||||: 1-1 |
DB 36 PAHVLQDDLT 45

RESULT 5
E84653
probable poly(A) polymerase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84653
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84653
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-414 <STO>
A:Cross-references: GB:AE002093; NID:g3643598; PIDN:AAC42245.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g25850
A:Map position: 2

Query Match 46.7%; Score 43; DB 2; Length 414;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 7 QDDVSFPW 15
:| |: :||| |
DB 295 EDDLSPFW 303

RESULT 6
A35087
arachidonate 12-lipoxygenase (EC 1.13.11.31) - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A35087
R:Yoshimoto, T.; Suzuki, H.; Yamamoto, S.; Takai, T.; Yokoyama, C.; Tanabe, T.
Proc. Natl. Acad. Sci. U.S.A. 87, 2142-2146, 1990
A:Title: Cloning and sequence analysis of the cDNA for arachidonate 12-lipoxygenase
A:Reference number: A35087; MUID:90192763; PMID:2315307
A:Accession: A35087
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-663 <YOS>
A:Cross-references: GB:M31417
C:Superfamily: arachidonate 5-lipoxygenase
C:Keywords: oxidoreductase

Query Match 45.7%; Score 42; DB 1; Length 663;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSFPW 15
|||||: 1-1 |
DB 59 HLLQDDAWFCNW 80

RESULT 7
T51500
hypothetical protein F5E19_20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51500
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
submitted to the Protein Sequence Database, August 2000
A:Reference number: Z25394
A:Accession: T51500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1280 <SAT>
A:Cross-references: EMBL:AL391147
A:Experimental source: cultivar Columbia; BAC clone F5E19
C:Genetics:
A:Map position: 5
A:Introns: 272/3; 301/2; 331/1; 425/1; 557/1; 808/2; 885/2; 927/1; 1019/3
A:Note: F5E19_20

Query Match 45.7%; Score 42; DB 2; Length 1280;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPA 14
 DB 765 IFSRLVEVSFPA 778
 :||:|:|:|:|:|:|

RESULT 8
 S39481
 glutamate-ammonia ligase (EC 6.3.1.2) 1-5, cytosolic - maize (fragment)
 N:Alternate names: glutamine synthetase
 C:Species: Zea mays (maize)
 C:Date: 07-Apr-1994 #sequence_revision 08-Nov-1996 #text_change 03-Jun-2002
 C:Accession: S39481
 R:Li, M.G.; Villemur, R.; Hussey, P.J.; Silflow, C.D.; Gantt, J.S.; Snustad, D.P.
 Plant Mol. Biol. 23, 401-407, 1993
 A:Title: Differential expression of six glutamine synthetase genes in Zea mays.
 A:Reference number: S39477; PMID:94033318; PMID:8106013
 A:Accession: S39481
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-245 <LIM>
 A:Cross-references: EMBL:X65930; NID:q434331; PIDN:CAA46723.1; PID:q1360700
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
 C:Superfamily: glutamate-ammonia ligase
 C:Keywords: cytosol; ligase

Query Match 45.1%; Score 41.5; DB 2; Length 245;
 Best Local Similarity 69.2%; Pred. No. 18;
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 5 LLQDDVSFP-AMP 16
 DB 22 LLQDVSFPLGWP 34
 |||||:|:|:|

RESULT 9
 S39477
 glutamate-ammonia ligase (EC 6.3.1.2) 1-1, cytosolic - maize
 N:Alternate names: glutamine synthetase
 C:Species: Zea mays (maize)
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 03-Jun-2002
 C:Accession: S39477
 R:Li, M.G.; Villemur, R.; Hussey, P.J.; Silflow, C.D.; Gantt, J.S.; Snustad, D.P.
 Plant Mol. Biol. 23, 401-407, 1993
 A:Title: Differential expression of six glutamine synthetase genes in Zea mays.
 A:Reference number: S39477; PMID:94033318; PMID:8106013
 A:Accession: S39477
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-357 <LIM>
 A:Cross-references: EMBL:X65926; NID:q434323; PIDN:CAA46719.1; PID:q434324
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992
 C:Superfamily: glutamate-ammonia ligase
 C:Keywords: cytosol; ligase

Query Match 45.1%; Score 41.5; DB 2; Length 357;
 Best Local Similarity 69.2%; Pred. No. 28;
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 5 LLQDDVSFP-AMP 16
 DB 134 LLQDVSFPLGWP 146
 |||||:|:|:|

RESULT 10
 H82573
 excinuclease ABC subunit C XF2311 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A70885

C:Accession: H82573
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: AB2515; PMID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: H82573
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-627 <SIM>
 A:Cross-references: GB:AE004042; GB:AE003849; NID:g9107469; PIDN:AAF85110.1; GSPDB:GI
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R
 Briones, M.R.P.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrei
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; F
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; L
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martini
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 A:Genetics:
 C:Gene: XF2311
 C:Superfamily: excinuclease ABC chain C

Query Match 45.1%; Score 41.5; DB 2; Length 627;
 Best Local Similarity 47.4%; Pred. No. 54;
 Matches 9; Conservative 2; Mismatches 1; Indels 7; Gaps 1;

QY 5 LLQDDVSFP-----ANP 16
 DB 105 LLRDDKSPYVLLTREAWP 123
 |||||:|:|

RESULT 11
 AB4104
 hypothetical protein BH3633 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: AB4104
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; F
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
 A:Reference number: AB3650; PMID:20512582; PMID:11058132
 A:Accession: AB4104
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <STO>
 A:Cross-references: GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BAB07352.1; GSPDB:C
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3633

Query Match 44.6%; Score 41; DB 2; Length 160;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 HLQDDVSF 12
 DB 50 HLQDDLSF 58
 ||:|:|:|

RESULT 12
 A70885
 hypothetical protein Rv279Ac - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70885

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: F82746
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-227 <COL>
 A:Cross-references: GB:AL008967; GB:AL123456; NID:g3261491; PIDN:CAAL1589.1; PID:g262431
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv2794c

Query Match 44.6%; Score 41; DB 2; Length 227;
 Best Local Similarity 46.7%; Pred. No. 20;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPWP 16
 |||:::|
 Db 70 PAPILKGDKEPCWP 84

RESULT 13

T21161

hypothetical protein F20E11.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T21161

R:Ainscough, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: F21161

A:Accession: T21161

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-338 <WIL>

A:Cross-references: EMBL:281508; PIDN:CAB04147.1; GSPDB:GN00023; CESP:F20E11.10

A:Experimental source: clone F20E11

C:Genetics:

A:Gene: CESP:F20E11.10

A:Map position: 5

A:Introns: 110/1; 173/3

C:Superfamily: Caenorhabditis hypothetical protein C49G7.2

Query Match 44.6%; Score 41; DB 2; Length 338;
 Best Local Similarity 50.0%; Pred. No. 32;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFP 13
 |||:::|
 Db 186 PVYIADDTVP 197

RESULT 14

F82746

DNA processing chain A XF0924 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: F82746

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10510347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: F82746

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-387 <SIM>

A:Cross-references: GB:AE003931; GB:AE003849; NID:g9105834; PIDN:AAF83734.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier
 as-Neco, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima J.P.; Krieger, J.E.; Kuramae, E.B.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 M.; Tsubako, M.H.; Vallada, H.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF0924

Query Match 44.6%; Score 41; DB 2; Length 387;
 Best Local Similarity 42.9%; Pred. No. 38;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHLQDDVSFPWP 16
 |||:::|
 Db 307 AHTLRERLDAPSWP 320

RESULT 15

T22856

hypothetical protein F57F5.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Jun-2002

C:Accession: T22856

R:Harris, B.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z19627

A:Accession: T22856

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-547 <WIL>

A:Cross-references: EMBL:Z75953; PIDN:CAB00101.1; GSPDB:GN00023; CESP:F57F5.5

A:Experimental source: clone F57F5

C:Genetics:

A:Gene: CESP:F57F5.5

A:Map position: 5

A:Introns: 43/1; 129/2; 166/3; 187/1; 289/3; 341/2; 450/3; 499/3

F:11-60/Domain: protein kinase C zinc-binding repeat homology <KZN1>

F:89-139/Domain: protein kinase C zinc-binding repeat homology <KZN2>

Query Match 44.6%; Score 41; DB 2; Length 547;
 Best Local Similarity 54.5%; Pred. No. 57;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPWP 15
 |||:::|
 Db 429 ILNDDVLYPVW 439

Search completed: May 14, 2003, 15:32:58

Job time : 46 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:24:22 ; Search time 11 seconds
(without alignments)
60.329 Million cell updates/sec

Title: US-10-016-725-16
Perfect score: 92
Sequence: 1 MPAHLQDDVFPAMP 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	60.9	359	1 ACOD_HUMAN	O00767 homo sapien
2	48	52.2	359	1 ACOD_SHEEP	O62849 ovis aries
3	46	50.0	359	1 ACOD_BOVIN	Q9tt94 bos taurus
4	45	48.9	1675	1 CLH1_HUMAN	Q00610 homo sapien
5	45	48.9	1675	1 CLH1_BOVIN	P49951 bos taurus
6	45	48.9	1675	1 CLH1_RAT	P11442 rattus norv
7	44	47.8	1687	1 VIT2_FUNHE	Q98893 fundulus he
8	42	45.7	662	1 LOX2_PIG	P16469 sus scrofa
9	41.5	45.1	357	1 GLN1_MAIZE	P38559 zea mays (m
10	41	44.6	707	1 KPC2_CAEEL	P34885 caenorhabdi
11	41	44.6	736	1 KPCE_RABIT	P10830 oryctolagus
12	41	44.6	737	1 KPCE_HUMAN	Q02156 homo sapien
13	41	44.6	737	1 KPCE_MOUSE	P16034 mus musculu
14	41	44.6	737	1 KPCE_RAT	P09216 rattus norv
15	41	44.6	743	1 KPC2_APLCA	Q16975 aplysia cal
16	41	44.6	847	1 MDOH_ECOLI	P33137 escherichia
17	40	43.5	243	1 IPT2_AGR75	P06524 agrobacteri
18	40	43.5	243	1 IPT2_AGR75	O04701 cynodon dac
19	40	43.5	246	1 MPCL_CYNDA	P49966 arabidopsis
20	40	43.5	495	1 SR53_ARATH	P49967 arabidopsis
21	40	43.5	495	1 SR53_ARATH	Q9yht4 brachydanio
22	40	43.5	583	1 HEMO_BRARE	P32337 saccharomyc
23	40	43.5	1089	1 IMB3_YEAST	O08892 cavia porce
24	39.5	42.9	389	1 5H1B_CAVPO	P03412 bovine leuk
25	39	42.4	308	1 ACOD_MESAU	Q64420 mesocricetu
26	39	42.4	354	1 ACOD_MESAU	Q60430 mesocricetu
27	39	42.4	355	1 NDF1_MESAU	Q13562 homo sapien
28	39	42.4	356	1 NDF1_HUMAN	Q60867 mus musculu
29	39	42.4	357	1 NDF1_MOUSE	O64289 rattus norv
30	39	42.4	357	1 NDF1_MOUSE	O70469 mus musculu
31	39	42.4	412	1 DOK2_MOUSE	Q59699 pseudomonas
32	39	42.4	495	1 HYDA_PSEPU	Q9rtt6 deinococcus
33	39	42.4	498	1 SYNC_DEIRA	

34	39	42.4	634	1 KPC3_DROME	P13678 drosophila
35	39	42.4	958	1 YGX7_YEAST	P53076 saccharomyc
36	39	42.4	1032	1 YGFK_ECO57	Q8xd75 escherichia
37	39	42.4	1032	1 YGFK_ECOLI	Q46811 escherichia
38	39	42.4	1293	1 ENTF_ECO57	Q8xbv9 escherichia
39	39	42.4	1293	1 ENTF_ECOLI	P11454 escherichia
40	39	42.4	1678	1 CLH1_DROME	P29742 drosophila
41	39	42.4	3432	1 POLG_JAEV1	P27395 j genome po
42	39	42.4	3432	1 POLG_JAEV5	P19110 j genome po
43	39	42.4	3432	1 POLG_JAEVJ	P32886 j genome po
44	39	42.4	5217	1 HTSL_COCCA	Q01886 cochliobolu
45	38.5	41.8	356	1 GLN2_PNAVU	P04771 phaseolus v

ALIGNMENTS

RESULT 1	ACOD_HUMAN	STANDARD;	PRT;	359 AA.
ID	ACOD_HUMAN	Q00767; Q16150; Q9Y695; Q9BS07; Q96KF6;		
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty acid desaturase) (Delta(9)-desaturase).			
GN	SCD.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RA	Al-Jeryan L., McCord A., Pierotti A.R., Craft J.A.;			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver, Brain, and Skin;			
RX	MEDLINE=99247918; PubMed=10229681;			
RA	Zhang L., Ge L., Parimoo S., Stenn K., Prouty S.M.;			
RT	"Human stearoyl-CoA desaturase: alternative transcripts generated from a single gene by usage of tandem polyadenylation sites.";			
RL	Biochem. J. 340:255-264(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Hoshino T., Ohtsu K.;			
RT	"Cloning, sequencing and expression of human stearoyl-CoA desaturase.";			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RA	Strausberg R.;			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE OF 3-239 FROM N.A.			
RC	TISSUE=Adipose tissue;			
RA	MEDLINE=94222609; PubMed=7909540;			
RX	Li J., Ding S.-F., Habib N.A., Ferner B.F., Wood C.B., Gilmour R.S.;			
RT	"Partial characterization of a cDNA for human stearoyl-CoA desaturase and changes in its mRNA expression in some normal and malignant tissues.";			
RL	Biochem. J. 0:0-0(2001).			
RN	[6]			
RP	Int. J. Cancer 57:348-352(1994).			
RN	[6]			
RP	SEQUENCE OF 1-9 FROM N.A.			
RA	Zhang L., Ge L., Tran T., Stenn K., Prouty S.M.;			
RT	"Isolation and characterization of the human stearoyl-CoA desaturase gene promoter: requirement of a conserved CCAAT cis-element.";			
RL	Biochem. J. 0:0-0(2001).			
RN	[6]			
RP	FUNCTION: TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEAROYL-COA DESATURASE SYSTEM, THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A			

CC SPECTRUM OF FATTY ACYL-CoA SUBSTRATES AMONG WHICH PALMITOYL-CoA
 CC AND STEAROYL-CoA.
 CC -!- CATALYTIC ACTIVITY: Stearoyl-CoA + AH(2) + O(2) = oleoyl-CoA + A +
 CC 2 H(2)O.
 CC -!- COFACTOR: IRON.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum (Probable).
 CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
 CC AND/OR BE INVOLVED IN METAL ION BINDING.
 CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 CC EMBL; Y13647; CAA73998.1; -.
 CC EMBL; AF097514; AAD29870.1; -.
 CC EMBL; AB032261; BAA93510.1; -.
 CC EMBL; BC005807; AAB05807.1; -.
 CC EMBL; S70284; AAB30631.1; -.
 CC EMBL; AF320307; AAK54510.1; -.
 CC Genew; HGNC:10571; SCD.
 CC MIM; 604031; -.
 CC InterPro; IPR001522; Desaturase.
 CC InterPro; IPR001225; FA_desaturase.
 CC Pfam; PF00487; FA_desaturase; 1.
 CC PRINTS; PR00075; FACDDSATRASE.
 CC ProDom; PD002221; Desaturase; 1.
 CC PROSITE; PS00476; FATTY_ACID_DESATUR_1; 1.
 CC Oxidoreductase; Fatty acid biosynthesis; Transmembrane;
 CC Endoplasmic reticulum; Iron.
 CC TRANSMEM 76 96 POTENTIAL.
 CC TRANSMEM 98 118 POTENTIAL.
 CC TRANSMEM 223 243 POTENTIAL.
 CC TRANSMEM 315 335 POTENTIAL.
 CC DOMAIN 120 125 HISTIDINE BOX-1.
 CC DOMAIN 157 161 HISTIDINE BOX-2.
 CC DOMAIN 298 302 HISTIDINE BOX-3.
 CC CONFLICT 5 5 L -> M (IN REF. 5).
 CC CONFLICT 8 8 D -> E (IN REF. 5).
 CC CONFLICT 25 26 SR -> PG (IN REF. 1 AND 5).
 CC CONFLICT 224 224 M -> L (IN REF. 2 AND 3).
 CC CONFLICT 237 237 F -> C (IN REF. 5).
 CC CONFLICT 320 320 T -> N (IN REF. 1).
 CC CONFLICT 326 326 C -> W (IN REF. 1).
 CC CONFLICT 333 333 A -> T (IN REF. 1).
 CC CONFLICT 356 359 MISSING (IN REF. 4).
 CC SEQUENCE 359 AA; 41522 MW; ED56A63BDB850F05 CRC64;
 CC
 CC Query Match 60.9%; Score 56; DB 1; Length 359;
 CC Best Local Similarity 90.9%; Pred. No. 0.041;
 CC Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1 MPAHLQDDVS 11
 CC |||||:|
 CC Db 1 MPAHLQDDIS 11
 CC
 CC RESULT 2
 CC ACOD_SHEEP STANDARD; PRT; 359 AA.
 CC ID ACOD_SHEEP STANDARD; PRT; 359 AA.
 CC AC 062849;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty
 CC acid desaturase) (delta(9)-desaturase).
 CC GN SCD.
 CC OS Ovis aries (Sheep).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Caprinae; Ovis.
 CC NCBI_TaxID=9940;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=Y1089; TISSUE=adipose tissue;
 CC RX MEDLINE=98223428; PubMed=9554990;
 CC RA Ward R.J., Travers M.T., Richards S.E., Vernon R.G., Salter A.M.,
 CC Buttery P.J., Barber M.C.;
 CC "Stearoyl-CoA desaturase mRNA is transcribed from a single gene in the
 CC ovine genome.";
 CC Biochim. Biophys. Acta 1391:145-156(1998).
 CC -!- FUNCTION: TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEAROYL-CoA
 CC DESATURASE SYSTEM, THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED
 CC CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A
 CC SPECTRUM OF FATTY ACYL-CoA SUBSTRATES AMONG WHICH PALMITOYL-CoA
 CC AND STEAROYL-CoA (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Stearoyl-CoA + AH(2) + O(2) = oleoyl-CoA + A +
 CC 2 H(2)O.
 CC -!- COFACTOR: IRON.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum (Probable).
 CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
 CC AND/OR BE INVOLVED IN METAL ION BINDING.
 CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 CC EMBL; AJ001048; CAA04502.1; -.
 CC InterPro; IPR001522; Desaturase.
 CC InterPro; IPR001225; FA_desaturase.
 CC Pfam; PF00487; FA_desaturase; 1.
 CC PRINTS; PR00075; FACDDSATRASE.
 CC ProDom; PD002221; Desaturase; 1.
 CC PROSITE; PS00476; FATTY_ACID_DESATUR_1; 1.
 CC Oxidoreductase; Fatty acid biosynthesis; Transmembrane;
 CC Endoplasmic reticulum; Iron.
 CC TRANSMEM 76 96 POTENTIAL.
 CC TRANSMEM 98 118 POTENTIAL.
 CC TRANSMEM 223 243 POTENTIAL.
 CC TRANSMEM 315 335 POTENTIAL.
 CC DOMAIN 120 125 HISTIDINE BOX-1.
 CC DOMAIN 157 161 HISTIDINE BOX-2.
 CC DOMAIN 298 302 HISTIDINE BOX-3.
 CC SEQUENCE 359 AA; 41671 MW; 219CFEBB1E353418 CRC64;
 CC
 CC Query Match 52.2%; Score 48; DB 1; Length 359;
 CC Best Local Similarity 72.7%; Pred. No. 0.95;
 CC Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1 MPAHLQDDVS 11
 CC |||||:|
 CC Db 1 MPAHLQDDIS 11
 CC
 CC RESULT 3
 CC ACOD_BOVIN STANDARD; PRT; 359 AA.
 CC ID ACOD_BOVIN STANDARD; PRT; 359 AA.
 CC AC Q9TT94;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty
 CC acid desaturase) (delta(9)-desaturase).
 CC GN SCD.
 CC OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Adipose tissue;
 RA Chung M.I., Ha S.H., Baik M.G., Choi Y.J.;
 RT "Cloning and characterization of full-coding cDNA of bovine stearoyl
 RT CoA desaturase from adipose tissue.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: TERMINAL COMPONENT OF THE LIVER MICROSOMAL STEARYL-COA
 CC DESATURASE SYSTEM. THAT UTILIZES O(2) AND ELECTRONS FROM REDUCED
 CC CYTOCHROME B(5) TO CATALYZE THE INSERTION OF A DOUBLE BOND INTO A
 CC SPECTRUM OF FATTY ACYL-COA SUBSTRATES AMONG WHICH PALMITOYL-COA
 CC AND STEAROYL-COA (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Stearoyl-CoA + AH(2) + O(2) = oleoyl-CoA + A +
 CC 2 H(2)O.
 CC -!- COFACTOR: IRON.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.. Endoplasmic
 CC reticulum (Probable).
 CC -!- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
 CC AND/OR BE INVOLVED IN METAL ION BINDING.
 CC -!- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF188710; RAF22305.1; -
 CC InterPro; IPR001522; Desaturase.
 CC InterPro; IPR001225; FA_desaturase.
 CC Pfam; PF00487; FA_desaturase; 1.
 CC PRINTS; PR00075; FACDSDSTRASE.
 CC ProDom; PD002221; Desaturase; 1.
 CC PROSITE; PS00476; FATTY_ACID_DESATUR_1; 1.
 CC Oxidoreductase; Fatty acid biosynthesis; Transmembrane;
 CC Endoplasmic reticulum; Iron.
 CC TRANSMEM 76 96 POTENTIAL.
 CC TRANSMEM 98 118 POTENTIAL.
 CC TRANSMEM 223 243 POTENTIAL.
 CC TRANSMEM 315 335 POTENTIAL.
 CC DOMAIN 120 125 HISTIDINE BOX-1.
 CC DOMAIN 157 161 HISTIDINE BOX-2.
 CC DOMAIN 298 302 HISTIDINE BOX-3.
 CC SEQUENCE 359 AA; 41755 MW; 22FB69FBD1846C33 CRC64;
 Query Match 50.0%; Score 46; DB 1; Length 359;
 Best Local Similarity 63.6%; Pred. No. 2.1;
 Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPAHLQDDVS 11
 DB 1 MPAHLQDEIS 11
 RESULT 4
 ID CLHL_HUMAN STANDARD; PRT; 1675 AA.
 AC Q00610;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Clathrin heavy chain 1 (CLH-17).
 GN CLH1 OR CLH17 OR KIAA0034.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.
 RP TISSUE=Bone marrow;
 RC MEDLINE=96051387; PubMed=7584026;
 RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
 RT "Prediction of the coding sequences of unidentified human genes. I.
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
 RT analysis of randomly sampled cDNA clones from human immature myeloid
 RT cell line KG-1.";
 RL DNA Res. 1:27-35(1994).
 RN [2]
 RP SEQUENCE OF 560-864 FROM N.A.
 RC TISSUE=Colon;
 RA Dodge G.R., Kowalsky I., McBride O.W., Yi H.F., Chu M.L., Saitta B.,
 RA Stokes D.G., Iozzo R.V.;
 RT "Human clathrin heavy chain (CLTC): partial molecular cloning,
 RT expression, and mapping of the gene to human chromosome 17q11-qter.";
 RL Genomics 11:174-178(1991).
 CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
 CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
 CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
 CC TRANS GOLGI NETWORK.
 CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
 CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
 CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
 CC PH AND THE CONCENTRATION OF CALCIUM.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
 CC VESICLES.
 CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
 CC -!- DATABASE: NAMP-Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CLTCD360.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D21260; BA04801.1; -
 CC EMBL; X55878; CA039363.1; -
 CC PIR; A40573; A40573.
 CC HSP; P11442; LBPO.
 CC Genew; HGNC:2092; CLTC.
 CC MIM; 118995; -
 CC InterPro; IPR001473; Clathrin_propel.
 CC InterPro; IPR000547; Clathrin_repeat.
 CC Pfam; PF00637; Clathrin; 7.
 CC Pfam; PF01394; Clathrin_propel; 7.
 CC SMART; SM00299; CLH; 7.
 CC Coated pits.
 CC DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.
 CC DOMAIN 480 523 FLEXIBLE LINKER.
 CC DOMAIN 524 1675 HEAVY CHAIN ARM.
 CC DOMAIN 524 634 DISTAL SEGMENT.
 CC DOMAIN 639 1675 PROXIMAL SEGMENT.
 CC DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,
 CC INVOLVED IN LATTICE DISASSEMBLY
 CC (POTENTIAL).
 CC BINDING 1213 1522 LIGHT CHAIN (BY SIMILARITY).
 CC DOMAIN 1550 1675 TRIMERIZATION (BY SIMILARITY).
 CC CONFLICT 560 560 Q -> R (IN REF. 2).
 CC CONFLICT 817 817 G -> V (IN REF. 2).
 CC SEQUENCE 1675 AA; 191614 MW; 6C4F2D54950079E2 CRC64;
 Query Match 48.9%; Score 45; DB 1; Length 1675;
 Best Local Similarity 53.3%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MPAHLQDDVSFPAP 15

```

Db 99 MKAHMTDDVTFWKW 113
| | | : | | | : |
RESULT 5
CLH_BOVIN STANDARD; PRT; 1675 AA.
AC P49951;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Clathrin heavy chain.
GN CLTC.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=96028100; PubMed=7585943;
RA Liu S.-H., Wong M.L., Craik C.S., Brodsky F.M.;
RT "Regulation of clathrin assembly and trimerization defined using
RT recombinant triskelion hubs.";
RL Cell 83:257-267(1995).
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC TRANS GOLGI NETWORK.
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
CC PH AND THE CONCENTRATION OF CALCIUM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -!- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF
CC THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
CC THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
CC CLATHRIN LATTICE.
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U31757; AAC48524.1; -
CC HSSP; P11442; LBPO.
CC InterPro; IPR001473; Clathrin_propel.
CC InterPro; IPR000547; Clathrin_repeat.
CC Pfam; PF00637; Clathrin; 7.
CC Pfam; PF01394; Clathrin_propel; 7.
CC SMART; SM00299; CLH; 7.
CC Coated pits.
CC FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.
CC FT DOMAIN 480 523 FLEXIBLE LINKER.
CC FT DOMAIN 524 1675 HEAVY CHAIN ARM.
CC FT DOMAIN 524 634 DISTAL SEGMENT.
CC FT DOMAIN 639 1675 PROXIMAL SEGMENT.
CC FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,
CC INVOLVED IN LATTICE DISASSEMBLY
CC (POTENTIAL).
CC FT BINDING 1213 1522 LIGHT CHAIN.
CC FT DOMAIN 1550 1675 TRIMERIZATION.
CC FT SEQUENCE 1675 AA; 191587 MW; 6C4F2D54801579E2 CRC64;
Query Match 48.9%; Score 45; DB 1; Length 1675;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 MPAHLQDDVSYFPAW 15
| | | : | | | : |
Db 99 MKAHMTDDVTFWKW 113
| | | : | | | : |
RESULT 6
CLH_RAT STANDARD; PRT; 1675 AA.
AC P11442;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Clathrin heavy chain.
GN CLTC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [-]
RP SEQUENCE FROM N.A.
RX MEDLINE=88097376; PubMed=3480512;
RA Kirchhausen T., Harrison S.C., Chow E.P., Mattaliano R.J.,
RA Ramchandran K.L., Smart J., Brosius J.;
RT "Clathrin heavy chain: molecular cloning and complete primary
RT structure.";
RL Cell 95:563-573(1998).
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC TRANS GOLGI NETWORK.
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
CC PH AND THE CONCENTRATION OF CALCIUM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J03583; AAA0874.1; -
CC PIR; A39941; LRTH.
CC PDB; LBPO; 06-APR-99.
CC InterPro; IPR001473; Clathrin_propel.
CC InterPro; IPR000547; Clathrin_repeat.
CC Pfam; PF00637; Clathrin; 7.
CC Pfam; PF01394; Clathrin_propel; 7.
CC SMART; SM00299; CLH; 7.
CC Coated pits; 3D-structure.
CC FT DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.
CC FT DOMAIN 480 523 FLEXIBLE LINKER.
CC FT DOMAIN 524 1675 HEAVY CHAIN ARM.
CC FT DOMAIN 524 634 DISTAL SEGMENT.
CC FT DOMAIN 639 1675 PROXIMAL SEGMENT.
CC FT DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,
CC INVOLVED IN LATTICE DISASSEMBLY
CC (POTENTIAL).
CC FT BINDING 1213 1522 LIGHT CHAIN (BY SIMILARITY).

```

FT	CARBOHYD	1093	1093	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1179	1179	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1257	1257	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1292	1292	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1342	1342	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1361	1361	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1366	1366	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1390	1390	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1577	1577	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1655	1655	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1687	AA; 186005	MM; 4965BB9DBF9428F	CRC64;

Query Match 47.8%; Score 44; DB 1; Length 1687;
Best Local Similarity 61.5%; Pred. No. 26;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY	2	PAHLIQDDVSPFA 14
DB	831	PAHLIKDSISKKA 843

RESULT 8
LOX2_PIG
ID LOX2_PIG STANDARD; PRT; 662 AA.
AC P16469;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Arachidonate 12-lipoxygenase, 12S-type (EC 1.13.11.31) (12-LOX).
GN ALOX12.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
EN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Leukocyte;
RX MEDLINE=90192763; PubMed=2315307;
RA Yoshimoto T., Suzuki H., Yamamoto S., Takai T., Yokoyama C.,
RA Tanabe T.;
RT "Cloning and sequence analysis of the cDNA for arachidonate 12-
RT lipoxygenase of porcine leukocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2142-2146(1990).
RN [2]
RP MUTAGENESIS.
RX MEDLINE=94137772; PubMed=8305485;
RA Suzuki H., Kishimoto K., Yoshimoto T., Yamamoto S., Kanai F.,
RA Ebina Y., Miyatake A., Tanabe T.;
RT "Site-directed mutagenesis studies on the iron-binding domain and the
RT determinant for the substrate oxygenation site of porcine leukocyte
RT arachidonate 12-lipoxygenase.";
RL Biochim. Biophys. Acta 1210:308-316(1994).
CC -1- FUNCTION: OXYGENASE AND 14,15-LEUKOTRIENE A4 SYNTHASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Arachidonate + O(2) = (5Z,8Z,10E,14Z)-(12S)-
CC 12-hydroperoxyicosaa-5,8,10,14-tetraenoate.
CC -1- COFACTOR: IRON (CONTAINS 0.45 ATOM OF IRON PER MOLECULE).
CC -1- PATHWAY: Leukotrienes biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: LEUKOCYTES, PITUITARY GLAND, LUNG AND IN
CC VERY SMALL AMOUNT IN JEJUNUM AND SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE LIPOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31417; AAA31068.1; -;
DR EMBL; D10621; BAA01471.1; -;

```

DR EMBL; D10616; BAA01471.1; JOINED.
DR EMBL; D10617; BAA01471.1; JOINED.
DR EMBL; D10618; BAA01471.1; JOINED.
DR EMBL; D10619; BAA01471.1; JOINED.
DR EMBL; D10620; BAA01471.1; JOINED.
DR F1R; A35087; A35087.
DR HSP; P12530; ILOX.
DR InterPro; IPR000907; Lipoxigenase.
DR InterPro; IPR001024; Lipoxigenase_LH2.
DR Pfam; PF00305; lipoxigenase; 1.
DR Pfam; PF01477; PLAT; 1.
DR PRINTS; PR00087; LIPOXYGENASE.
DR SMART; SM00308; LH2; 1.
DR PROSITE; PS00081; LIPOXYGENASE_2; 1.
DR PROSITE; PS00711; LIPOXYGENASE_1; 1.
KW Oxidoreductase; Dioxigenase; Iron; Leukotriene biosynthesis.
FT INIT_MET 0
FT METAL 360 360 IRON (BY SIMILARITY).
FT METAL 365 365 IRON (BY SIMILARITY).
FT METAL 540 540 IRON (BY SIMILARITY).
FT METAL 662 662 IRON (BY SIMILARITY).
SQ SEQUENCE 662 AA; 74912 MW; F1ABE1B20E368183 CRC64;

Query Match 45.7%; Score 42; DB 1; Length 662;
Best Local Similarity 66.7%; Pred. NO. 20;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 HLLQDDVSFP 15
DB 68 HLLQDDAWFCNW 79

RESULT 9
GLN1_MAIZE
ID GLN1_MAIZE STANDARD; PRT; 357 AA.
AC P38559;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase root isozyme 1 (EC 6.3.1.2) (Glutamate--ammonia
DE ligase) (GSI22).
GN GLN6 OR GSI-1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. A188; TISSUE=Seedling;
RX MEDLINE=94033318; PubMed=8106013;
RA Li M.-G., Villemur R., Hussey P.J., Silflow C.D., Gantt J.S.,
RA Snustad D.P.;
RT "Differential expression of six glutamine synthetase genes in Zea
RT mays.";
RL Plant Mol. Biol. 23:401-407(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Golden cross Bantam T51; TISSUE=Leaf;
RA Sakakibara H., Kawabata S., Takahashi H., Hase T., Sugiyama T.;
RT "Molecular cloning of the family of glutamine synthetase genes from
RT maize: expression of genes for glutamine synthetase and ferredoxin-
RT dependent glutamate synthase in photosynthetic and non-photosynthetic
RT tissues.";
RL Plant Cell Physiol. 33:49-58(1992).
RN [3]
RP REVISIONS.
RA Sakakibara H.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS A ROLE IN THE FLOW OF NITROGEN INTO NITROGENOUS
CC ORGANIC COMPOUNDS.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) -> ADP + phosphate +
CC L-glutamine.

```

```

CC -!- SUBUNIT: HOMOOCTAMER.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: FOUND MAINLY IN THE CORTICAL TISSUES OF
CC SEEDLING ROOTS, AND IN THE ROOT TIP.
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65926; CAA46719.1; -
DR EMBL; D14579; BAA03433.1; -
DR F1R; S39477; S39477.
DR MaizeDB; 17151; -
DR InterPro; IPR001691; GLN_synth.
DR InterPro; IPR001637; GLNAdenyltn.
DR Pfam; PF00120; gln-synt; 1.
DR PROSITE; PS00180; GLNA_1; 1.
DR PROSITE; PS00181; GLNA_ATP; 1.
KW Ligase; Multigene family.
FT CONFLICT 48
FT CONFLICT 48 I -> S (IN REF. 2).
SQ SEQUENCE 357 AA; 39250 MW; 912A5E3BAF9CC2B8 CRC64;

Query Match 45.1%; Score 41.5; DB 1; Length 357;
Best Local Similarity 69.2%; Pred. NO. 12;
Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 5 LLQDDVSFP-AMP 16
DB 134 LLQDDVSNPLGWP 146

RESULT 10
KPC2_CAEEL
ID KPC2_CAEEL STANDARD; PRT; 707 AA.
AC P34885;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein kinase C-like 2 (EC 2.7.1.1-) (PKC1B).
GN KIN-13 OR PKC1B.
OS Caenorhabditis elegans
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94179345; PubMed=8132661;
RA Land M., Islas-Trejo A., Freedman J.H., Rubin C.S.;
RT "Structure and expression of a novel, neuronal protein kinase C
RT (PKC1B) from Caenorhabditis elegans. PKC1B is expressed selectively
RT in neurons that receive, transmit, and process environmental
RT signals.";
RL J. Biol. Chem. 269:9234-9244(1994).
CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH MEMBRANES AND THE
CC CYTOSKELETON.
CC -!- TISSUE SPECIFICITY: EXPRESSED SELECTIVELY IN NEURONS THAT RECEIVE,
CC TRANSMIT AND PROCESS ENVIRONMENTAL SIGNALS.
CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -!- PKC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U00181; AAA18259.1; -
 CC EMBL; U00965; AAA17996.1; -
 CC PIR; A53530; A53530.
 CC HSP; P28867; IPTQ.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR002219; DAG_PE-bind.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR000961; Pkinase_C.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00130; DAG_PE-bind; 2.
 CC Pfam; PF00168; C2; 1.
 CC Pfam; PF00433; pkinase_C; 1.
 CC PRINTS; PR000008; DAGPEDOMAIN.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00109; C1; 2.
 CC SMART; SM00239; C2; 1.
 CC SMART; SM00133; S_TK_X; 1.
 CC SMART; SM00220; S_TK_X; 1.
 CC PROSITE; PS00004; C2_DOMAIN_2; 1.
 CC PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
 CC PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC Transferase; Serine/threonine-protein kinase; ATP-binding; Zinc;
 KW Phorbol-ester binding; Repeat; Membrane; Cytoskeleton.
 FT DOMAIN 1 98
 FT DOMAIN 171 221
 FT DOMAIN 249 298
 FT DOMAIN 378 638
 FT NP_BIND 384 392
 FT BINDING 407 407
 FT ACT_SITE 502 502
 FT ACT_SITE 502 502
 SQ SEQUENCE 707 AA; 3DC762C8A7A7BA64 CRC64;

Query Match
 Best Local Similarity 44.6%; Score 41; DB 1; Length 707;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLOQDVGFPAW 15
 Db 589 ILNDDVLYPVW 599

RESULT 11

KPCE_RABIT
 ID KPCE_RABIT STANDARD; PRT; 736 AA.
 AC P10830;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Protein kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).
 GN PKCE.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 RN NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88223367; PubMed=3370672;
 RA Omo S., Akita Y., Konno Y., Imajoh S., Suzuki K.;
 RT "A novel phorbol ester receptor/protein kinase, nPKC, distantly
 RT related to the protein kinase C family";
 RL Cell 53:731-741(1988).
 CC -!- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,

CC SERINE- AND THREONINE-SPECIFIC ENZYME.
 CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
 CC -!- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PKC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M20014; AAA31426.1; -
 CC PIR; A29880; KIRBCE.
 CC HSP; P28867; IPTQ.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR002219; DAG_PE-bind.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR000961; Pkinase_C.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00130; DAG_PE-bind; 2.
 CC Pfam; PF00168; C2; 1.
 CC Pfam; PF00433; pkinase_C; 1.
 CC PRINTS; PR000008; DAGPEDOMAIN.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00109; C1; 2.
 CC SMART; SM00239; C2; 1.
 CC SMART; SM00133; S_TK_X; 1.
 CC SMART; SM00220; S_TK_X; 1.
 CC PROSITE; PS00004; C2_DOMAIN_2; 1.
 CC PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 CC PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC Repeat; ATP-binding; Transferase; Phosphorylation;
 KW Serine/threonine-protein kinase; Phorbol-ester binding; Zinc.
 FT DOMAIN 1 99
 FT DOMAIN 170 220
 FT DOMAIN 243 292
 FT DOMAIN 407 667
 FT NP_BIND 413 421
 FT BINDING 436 436
 FT ACT_SITE 531 531
 FT MOD_RES 702 702
 FT MOD_RES 709 709
 SQ SEQUENCE 736 AA; 83515 MW; 261C4FE59E9BEE CRC64;

Query Match
 Best Local Similarity 44.6%; Score 41; DB 1; Length 736;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLOQDVGFPAW 15
 Db 618 ILNDDVLYPVW 628

RESULT 12

KPCE_HUMAN
 ID KPCE_HUMAN STANDARD; PRT; 737 AA.
 AC Q02156; Q9UE81;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).
 GN PKCE OR PKCE.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93003318; PubMed=1382605;
 RA Basta P., Strickland M.B., Holmes W., Loomis C.R., Ballas L.M.,
 RA Burns D.J.;
 RT "Sequence and expression of human protein kinase C-epsilon.";
 RL Biochim. Biophys. Acta 1132:154-160(1992).
 RN [2]
 RP SEQUENCE OF 1-116 FROM N.A.
 RA Waterston R.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.
 CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
 CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PKC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X65293; CAA46388.1; -;
 CC EMBL; U51244; AAD08855.1; -;
 CC PIR; S28942; S28942.
 CC DR HSP; P28867; IPTQ.
 CC DR Genew; HGNC:9401; PRKCE.
 CC DR MIM; 176975; -;
 CC DR InterPro; IPR000008; C2.
 CC DR InterPro; IPR002219; DAG_PE-bind.
 CC DR InterPro; IPR000713; Euk_pkinase.
 CC DR InterPro; IPR000961; Pkinase_C.
 CC DR InterPro; IPR002290; Ser_thr_pkinase.
 CC DR Pfam; PF00069; pkinase; 1.
 CC DR Pfam; PF00130; DAG_PE-bind; 2.
 CC DR Pfam; PF00168; C2; 1.
 CC DR Pfam; PF00433; pkinase_C; 1.
 CC DR PRINTS; PR00008; DAGPEDOMAIN.
 CC DR ProDom; PD000001; Euk_pkinase; 1.
 CC DR SMART; SM00109; C1; 2.
 CC DR SMART; SM00239; C2; 1.
 CC DR SMART; SM00133; S_TK_X; 1.
 CC DR SMART; SM00220; S_TKC; 1.
 CC DR PROSITE; PS00004; C2_DOMAIN_2; 1.
 CC DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
 CC DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
 CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 CC DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC Repeat; ATP-binding; Transferase; Phosphorylation;
 CC Serine/threonine-protein kinase; Phorbol-ester binding; Zinc.
 CC DOMAIN 1 99 C2 DOMAIN.
 CC FT DOMAIN 170 220 PHORBOL-ESTER AND DAG BINDING 1.
 CC FT DOMAIN 243 292 PHORBOL-ESTER AND DAG BINDING 2.
 CC FT DOMAIN 408 668 PROTEIN KINASE.
 CC FT NP_BIND 414 422 ATP (BY SIMILARITY).
 CC FT BINDING 437 437 ATP (BY SIMILARITY).
 CC FT ACT_SITE 532 532 BY SIMILARITY.
 CC FT MOD_RES 703 703 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 CC FT MOD_RES 710 710 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 CC SEQUENCE 737 AA; 83673 MW; 85032D0A091A1F7F CRG64;

Query Match 44.6%; Score 41; DB 1; Length 737;
 Best Local Similarity 54.5%; Pred. No. 33;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 5 LIQDDVSFFAW 15
 Db 619 ILHDDVLYPFW 629
 RESULT 13
 KCPE_MOUSE
 ID KCPE_MOUSE STANDARD; PRT; 737 AA.
 AC P16054;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein Kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).
 GN PRKE OR PKCE OR PKCEA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89137541; PubMed=2917656;
 RA Schaap D., Parker P.J., Bristol A., Kriz R., Knopf J.;
 RT "Unique substrate specificity and regulatory properties of
 RT PKC-epsilon: a rationale for diversity.";
 RL FEBS Lett. 243:351-357(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RX MEDLINE=98127436; PubMed=9467942;
 RA Wang Q.J., ACS P., Goodnight J., Blumberg P.M., Mischak H.,
 RA Mushinski J.F.;
 RT "The catalytic domain of PKC-epsilon, in reciprocal PKC-delta and -
 RT epsilon chimeras, is responsible for conferring tumorigenicity to
 RT NIH3T3 cells, whereas both regulatory and catalytic domains of
 RT PKC-epsilon contribute to in vitro transformation.";
 RL Oncogene 16:53-60(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Wheeler D.L.;
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
 CC SERINE- AND THREONINE-SPECIFIC ENZYME.
 CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
 CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
 CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
 CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PKC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF028009; AAB84189.1; -;
 CC EMBL; AF325507; AAG53692.1; -;
 CC PIR; S02270; KIMSCF.
 CC HSP; P28867; IPTQ.
 CC MGI; MGI:97599; Pkce.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR002219; DAG_PE-bind.
 CC InterPro; IPR000713; Euk_pkinase.

```
DR InterPro; IPR000961; Pkinase.C.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00130; DAG_PE-bind; 2.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00433; pkinase.C; 1.
DR PRINTS; PR00008; DAGPEDOMAIN.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00109; C1; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00133; S_TK_X; 1.
DR SMART; SM00220; S_TK_X; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Repeat; ATP-binding; Transferase; phosphorylation;
KW Serine/threonine-protein kinase; phorbol-ester binding; zinc.
FT DOMAIN 1 99
FT DOMAIN 170 220 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 243 292 PHORBOL-ESTER AND DAG BINDING 2.
FT NP_BIND 414 422 ATP (BY SIMILARITY).
FT BINDING 437 437 ATP (BY SIMILARITY).
FT ACT_SITE 532 532 BY SIMILARITY.
FT MOD_RES 703 703 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 710 710 PHOSPHORYLATION (AUTO-) (POTENTIAL).
SQ SEQUENCE 737 AA; 83560 MW; 7AEBB8C10C99F57 CRC64;

Query Match 44.6%; Score 41; DB 1; Length 737;
Best Local Similarity 54.5%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLQDDVSFPW 15
Db 619 ILHDDVLYPVW 629
:| ||| :|

RESULT 14
KPC2_RAT
ID KPC2_RAT STANDARD; PRT; 737 AA.
AC P09216;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Protein kinase C, epsilon type (EC 2.7.1.-) (nPKC-epsilon).
GN PKCE OR PKCE.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1] NCBI_TaxID=10116;
RN TISSUE=Brain;
RP SEQUENCE FROM N.A.
RX MEDLINE=88198270; PubMed=2834397;
RA Ono Y., Fujii T., Ogita K., Kikkawa U., Igarashi K., Nishizuka Y.;
RT "The structure, expression, and properties of additional members of
RT the protein kinase C family.";
RL J. Biol. Chem. 263:6927-6932(1988).
[2]
RP SEQUENCE OF 135-297 FROM N.A.
RX MEDLINE=88083621; PubMed=3691811;
RA Ono Y., Fujii T., Ogita K., Kikkawa U., Igarashi K., Nishizuka Y.;
RT "Identification of three additional members of rat protein kinase C
RT family: delta-, epsilon- and zeta-subspecies.";
RL FEBS Lett. 226:125-128(1987).
CC -!- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
```

RN SEQUENCE FROM N.A.
RX MEDLINE-91332620; PubMed-1869917;
RA Kruger K.E., Sossin W.S., Sacktor T.C., Bergold P.J., Beushausen S.,
RA Schwartz J.H.;
RT "Cloning and characterization of Ca(2+)-dependent and Ca(2+)-
RT independent PKCs expressed in Aplysia sensory cells.";
RL J. Neurosci. 11:2303-2313(1991).
RN [2]
RX CHARACTERIZATION.
RX MEDLINE-93194877; PubMed-8449941;
RA Sossin W.S., Diaz-Arriastia R., Schwartz J.H.;
RT "Characterization of two isoforms of protein kinase C in the nervous
RT system of Aplysia californica.";
RL J. Biol. Chem. 268:5763-5768(1993).
RN [3]
RX DOMAIN C2.
RX MEDLINE-93348616; PubMed-8346555;
RA Sossin W.S., Schwartz J.H.;
RT "Ca(2+)-independent protein kinase Cs contain an amino-terminal domain
RT similar to the C2 consensus sequence.";
RL Trends Biochem. Sci. 18:207-208(1993).
RN [4]
RX CHARACTERIZATION.
RX MEDLINE-98334636; PubMed-9668085;
RA Pepio A.M., Fan X., Sossin W.S.;
RT "The role of C2 domains in Ca2+-activated and Ca2+-independent protein
RT kinase Cs in Aplysia.";
RL J. Biol. Chem. 273:19040-19048(1998).
RN [5]
RX ERRATUM.
RA Pepio A.M., Fan X., Sossin W.S.;
RL J. Biol. Chem. 273:22856-22856(1998).
RN [6]
RX CHARACTERIZATION.
RX MEDLINE-98138438; PubMed-9477951;
RA Pepio A.M., Sossin W.S.;
RT "The C2 domain of the Ca(2+)-independent protein kinase C Apl II
RT inhibits phorbol ester binding to the C1 domain in a phosphatidic
RT acid-sensitive manner.";
RL Biochemistry 37:1256-1263(1998).
CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,
CC SERINE- AND THREONINE-SPECIFIC ENZYME.
CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS (BY
CC SIMILARITY).
CC -1- ENZYME REGULATION: REQUIRES HIGH LEVELS OF PHOSPHATIDYL SERINE TO
CC BE ACTIVATED. THE PRESENCE OF THE C2 DOMAIN LOWERS THE AFFINITY OF
CC PROTEIN KINASE C ACTIVATORS FOR THE C1 DOMAINS AND THIS INHIBITION
CC CAN BE REMOVED BY PHOSPHATIDYL SERINE. PHOSPHATIDIC ACID, HOWEVER,
CC IS MUCH MORE POTENT THAN PHOSPHATIDYL SERINE IN REDUCING C2 DOMAIN-
CC MEDIATED INHIBITION, SUGGESTING THAT PHOSPHATIDYL SERINE MAY BE A
CC REQUIRED COFACTOR FOR THE ACTIVATION OF APL II.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NERVOUS TISSUES, OVOTESTIS AND
CC GUT.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M94884; AAA27771.1; -
DR HSP; P28867; LPTQ.

DR InterPro; IPR000008; C2.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000961; Pkinase_C.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00130; DAG_PE-bind; 2.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00433; pkinase_C; 1.
DR PRINTS; PR00008; DAGPEDOMAIN.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00109; C1; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00133; S_TK_X; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; FALSE_NEG.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; transferase; Serine/threonine-protein kinase;
KW Phorbol-ester binding; zinc; Repeat.
FT DOMAIN 1 108 C2 DOMAIN.
FT DOMAIN 177 226 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 248 297 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 403 663 PROTEIN KINASE.
FT NP_BIND 409 417 ATP (BY SIMILARITY).
FT BINDING 432 432 ATP (BY SIMILARITY).
FT ACT_SITE 527 527 BY SIMILARITY.
SQ SEQUENCE 743 AA; 84413 MW; 4C982C563CA2B659 CRC64;

Query Match 44.6%; Score 41; DB 1; Length 743;
Best Local Similarity 54.5%; Pred. No. 34;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 LQDDVSFPAP 15
:| ||| :| |
Db 614 ILHDDVLPVW 624

Search completed: May 14, 2003, 15:31:32
Job time : 17 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2003, 15:22:26 ; Search time 29 Seconds
(without alignments)
113.681 Million cell updates/sec

Title: US-10-016-725-16

Perfect score: 92

Sequence: 1 MPAHLQDDVSFPAMP 16

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	49	53.3	711	2 Q9AI02	Q9AI02 photorhabdu
2	48	52.2	359	6 Q95MI7	Q95MI7 capra hircu
3	48	52.2	359	6 Q9BG81	Q9BG81 capra hircu
4	48	52.2	359	6 Q8SQ76	Q8SQ76 bos taurus
5	47	51.1	688	16 Q984P3	Q984P3 rhizobium l
6	45	48.9	180	13 Q8UUR0	Q8UUR0 gallus galli
7	45	48.9	1675	13 Q8UUR1	Q8UUR1 gallus galli
8	44	47.8	68	5 Q9BPJ7	Q9BPJ7 conus texti
9	44	47.8	506	5 Q9GR90	Q9GR90 polyanthroca
10	43.5	47.3	518	10 Q81301	Q81301 arabidopsis
11	43	46.7	136	2 Q49796	Q49796 mycobacteri
12	43	46.7	208	4 Q9H718	Q9H718 homo sapien
13	43	46.7	437	16 Q99Z61	Q99Z61 streptococc
14	43	46.7	674	16 Q92QX2	Q92QX2 rhizobium m
15	43	46.7	753	10 Q8S864	Q8S864 oryza sativ
16	43	46.7	795	10 Q9LXX0	Q9LXX0 arabidopsis

17	43	46.7	800	10	Q82312	Q82312 arabidopsis
18	43	46.7	1725	13	Q8UW61	Q8UW61 oryzias lat
19	43	46.7	1793	5	Q9NKJ4	Q9NKJ4 leishmania
20	42.5	46.2	276	16	Q9KYM9	Q9KYM9 streptomyce
21	42	45.7	79	4	Q12938	Q12938 homo sapien
22	42	45.7	255	10	Q8VYH8	Q8VYH8 vernicia fo
23	42	45.7	270	16	Q9KY29	Q9KY29 streptomyce
24	42	45.7	551	10	Q94FS6	Q94FS6 linum usita
25	42	45.7	847	16	Q8YOC4	Q8YOC4 ralstonia s
26	42	45.7	1280	10	Q9LFP9	Q9LFP9 arabidopsis
27	42	45.7	2380	16	Q8XVE8	Q8XVE8 ralstonia s
28	41.5	45.1	388	2	Q9WVW9	Q9WVW9 streptomyce
29	41.5	45.1	627	16	Q9PB34	Q9PB34 xyella fas
30	41.5	45.1	1035	10	Q8S389	Q8S389 brassica na
31	41	44.6	160	16	Q9K6U2	Q9K6U2 bacillus ha
32	41	44.6	227	16	Q33336	Q33336 mycobacteri
33	41	44.6	240	11	Q63432	Q63432 rattus norv
34	41	44.6	338	5	Q9XV78	Q9XV78 caenorhabdi
35	41	44.6	357	13	Q9YGM2	Q9YGM2 gallus gall
36	41	44.6	387	16	Q9PEV4	Q9PEV4 xyella fas
37	41	44.6	490	13	Q42492	Q42492 fugu rubrip
38	41	44.6	567	4	Q8TEN4	Q8TEN4 homo sapien
39	41	44.6	660	16	Q8ZP78	Q8ZP78 salmonella
40	41	44.6	660	16	Q8Z795	Q8Z795 salmonella
41	41	44.6	681	16	Q8UGA9	Q8UGA9 agrobacteri
42	41	44.6	685	16	Q8YGA9	Q8YGA9 brucella me
43	41	44.6	707	5	Q20953	Q20953 caenorhabdi
44	41	44.6	772	6	Q95JG9	Q95JG9 sus scrofa
45	41	44.6	847	16	Q8ZQ26	Q8ZQ26 salmonella

ALIGNMENTS

RESULT 1

Q9AI02 ID Q9AI02 PRELIMINARY; PRT; 711 AA.
AC Q9AI02;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NGRD (Fragment).
GN NGRD.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photorhabdus.
OX NCBI_taxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC19;
RX MEDLINE=21225535; PubMed=11325940;
RA Cliche T.A., Bintrim S.B., Horswill A.R., Emsign J.C.;
RT "A Phosphopantetheinyl Transferase Homolog Is Essential for
RT Photorhabdus luminescens To Support Growth and Reproduction of the
RT Entomopathogenic Nematode Heterorhabditis bacteriophora.";
RL J. Bacteriol. 183:3117-3126(2001).
DR EMBL; AF288082; AAKI6083.1;
DR InterPro; IPR003838; DUF214.
DR Pfam; PF02687; DUF214; 1.
ET NON_TER 711 711
SQ SEQUENCE 711 AA; 79781 MW; 523E5CB4546CD805 CRC64;

Query Match 53.3%; Score 49; DB 2; Length 711;
Best Local Similarity 46.7%; Pred.No. 6.8;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPAMP 16

Db 364 PSRLVLDIIAPIWP 378

RESULT 2

Q95MI7

```

ID Q95MI7 PRELIMINARY; PRT; 359 AA.
AC Q95MI7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Stearoyl coenzyme A desaturase (EC 1.14.99.5).
GN SCD.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RA Yahyaoui M.H., Sanchez A., Folch J.M.;
RT "Nucleotide sequence of the goat stearoyl coenzyme A desaturase cDNA
and gene structure."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yahyaoui M.H., Sanchez A., Folch J.M.;
RT "Partial nucleotide sequence of the goat stearoyl coenzyme A
desaturase cDNA and gene structure."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF339909; AAK61862.1; -.
DR EMBL; AF422171; AAL29305.1; -.
DR EMBL; AF422166; AAL29305.1; JOINED.
DR EMBL; AF422167; AAL29305.1; JOINED.
DR EMBL; AF422168; AAL29305.1; JOINED.
DR EMBL; AF422169; AAL29305.1; JOINED.
DR EMBL; AF422170; AAL29305.1; JOINED.
DR InterPro; IPR001522; Desaturase.
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR PRODOM; PD002221; Desaturase; 1.
DR PROSITE; PS00476; FATTY_ACID_DESATUR_1; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 359 AA; 41586 MW; 052E5BE2B4463F88 CRC64;

Query Match 52.2%; Score 48; DB 6; Length 359;
Best Local Similarity 72.7%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
DB 1 MPAHLQEEIS 11

RESULT 3
Q9BG81 PRELIMINARY; PRT; 359 AA.
AC Q9BG81;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Acyl-CoA desaturase (EC 1.14.99.5) (Stearoyl-CoA desaturase) (Fatty
acid desaturase) (Delta(9)-desaturase).
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RA Bernard L., Leroux C., Hayes H., Gautier M., Chilliard Y., Martin P.;
RT "Genomic organization, chromosomal localization, and the complete 5 kb
cDNA sequence of the caprine SCD gene."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: STEAROYL-COA + AH(2) + O(2) = OLEOYL-COA + A +
2 H(2)O.
CC -1- COFACTOR: IRON (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC

```

```

CC RETICULUM (BY SIMILARITY).
CC -1- DOMAIN: THE HISTIDINE BOX DOMAINS MAY CONTAIN THE ACTIVE SITE
CC AND/OR BE INVOLVED IN METAL ION BINDING (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FATTY ACID DESATURASE FAMILY.
DR EMBL; AF325499; AAK01666.1; -.
DR InterPro; IPR001522; Desaturase.
DR InterPro; IPR001225; FA_desaturase.
DR Pfam; PF00487; FA_desaturase; 1.
DR PRINTS; PR00075; FACDDSATRASE.
DR PRODOM; PD002221; Desaturase; 1.
DR PROSITE; PS00476; FATTY_ACID_DESATUR_1; 1.
KW Endoplasmic reticulum; Fatty acid biosynthesis; Iron; Oxidoreductase;
KW Transmembrane.
SQ SEQUENCE 359 AA; 41582 MW; D27E5D55B44630CA CRC64;

Query Match 52.2%; Score 48; DB 6; Length 359;
Best Local Similarity 72.7%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
DB 1 MPAHLQEEIS 11

RESULT 4
Q8SQ76 PRELIMINARY; PRT; 359 AA.
AC Q8SQ76;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Stearoyl-CoA desaturase (EC 1.14.99.5).
GN SCD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RA Glimm D., Dong F., Kennelly J.;
RT "Bovine stearoyl-CoA desaturase gene structure and large scale SNP
analysis."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF481919; AAL99940.1; -.
DR EMBL; AF481915; AAL99940.1; JOINED.
DR EMBL; AF481916; AAL99940.1; JOINED.
DR EMBL; AF481917; AAL99940.1; JOINED.
DR EMBL; AF481918; AAL99940.1; JOINED.
KW Oxidoreductase.
SQ SEQUENCE 359 AA; 41733 MW; C7C9F6F52A90DD15 CRC64;

Query Match 52.2%; Score 48; DB 6; Length 359;
Best Local Similarity 72.7%; Pred. No. 4.9;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAHLQDDVS 11
DB 1 MPAHLQEEIS 11

RESULT 5
Q984P3 PRELIMINARY; PRT; 688 AA.
AC Q984P3;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Excinuclease ABC subunit C.
GN MLE7912.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.

```

```

OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003012; BAB54270.1; -
DR InterPro; IPR000445; HhH.
DR InterPro; IPR003583; HH1.1.
DR InterPro; IPR001943; UvrB/C.
DR InterPro; IPR004791; UvrC.
DR InterPro; IPR001162; UvrC_C.
DR InterPro; IPR000305; UvrC_N.
DR Pfam; PF01541; Excl_endo_N; 1.
DR Pfam; PF02151; Uvr; 1.
DR ProDom; PD005870; UvrC_C; 1.
DR SMART; SM00465; GYX; 1.
DR SMART; SM00278; HH1; 2.
DR TIGRFAMs; TIGR00194; uvrC; 1.
DR Complete proteome.
KW SEQUENCE 688 AA; 76147 MW; B10906D706614646 CRC64;

Query Match 51.1%; Score 47; DB 16; Length 688;
Best Local Similarity 61.5%; Pred. No. 14;
Matches 8; Conservative 3; Mismatches 0; Indels 2; Gaps 1;

QY 6 LQDDV--SFPWP 16
Db 529 IEDDJSFPWP 541
::||: |||||

RESULT 6
Q8UUR0 PRELIMINARY; PRT; 180 AA.
AC Q8UUR0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Clathrin heavy-chain (Fragment).
GN CHC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RA Wetley F.R., Hawkins S.F., Stewart A., Luzio J.P., Howard J.C.,
RA Jackson A.P.;
RT *Controlled elimination of clathrin heavy-chain expression in DT40
RT lymphocytes inhibits endocytosis but not lysosome biogenesis."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ429072; CAD2057.1; -
DR InterPro; IPR001473; Clathrin_prope.
DR InterPro; IPR000547; Clathrin_repeat.
DR Pfam; PF01394; Clathrin; 7.
DR SMART; SM00299; CLH; 7.
FT NON_TER 1
FT SEQUENCE 180 AA; 20155 MW; 85BC945FA30198E3 CRC64;

Query Match 48.9%; Score 45; DB 13; Length 180;
Best Local Similarity 53.3%; Pred. No. 7.6;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 15
Db 85 MKAHTMTDDVTFWKW 99
|||: |||||

RESULT 7
Q8UUR1 PRELIMINARY; PRT; 1675 AA.
AC Q8UUR1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Clathrin heavy-chain (Fragment).
GN CHC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RA Wetley F.R., Hawkins S.F., Stewart A., Luzio J.P., Howard J.C.,
RA Jackson A.P.;
RT *Controlled elimination of clathrin heavy-chain expression in DT40
RT lymphocytes inhibits endocytosis but not lysosome biogenesis."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ429072; CAD2057.1; -
DR InterPro; IPR001473; Clathrin_prope.
DR InterPro; IPR000547; Clathrin_repeat.
DR Pfam; PF01394; Clathrin; 7.
DR SMART; SM00299; CLH; 7.
FT NON_TER 1
FT SEQUENCE 1675 AA; 191599 MW; 337BE0B5423592DA CRC64;

Query Match 48.9%; Score 45; DB 13; Length 1675;
Best Local Similarity 53.3%; Pred. No. 81;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPAHLQDDVSFPWP 15
Db 99 MKAHTMTDDVTFWKW 113
|||: |||||

RESULT 8
Q8BPJ7 PRELIMINARY; PRT; 68 AA.
AC Q8BPJ7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Conotoxin scaffold III/IV.
OS Conus textile (Cloth-of-gold cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=6494;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21105969; PubMed=11158371;
RA Conticello S.G., Gilad Y., Avidan N., Ben-Asher E., Levy Z.,
RA Fainzilber M.;
RT *Mechanisms for Evolving Hypervariability: The Case of Conopeptides."
RL Mol. Biol. Evol. 18:120-131(2001).
DR EMBL; AF214926; AAG60354.1; -
DR SEQUENCE 68 AA; 7603 MW; 54CA322FFA7CB1B0 CRC64;

Query Match 47.8%; Score 44; DB 5; Length 68;
Best Local Similarity 53.3%; Pred. No. 4;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPWP 16
Db 31 PAERMQDDISFEQHP 45
|||: |||||

```

RESULT 9

Q9GR90
ID Q9GR90 PRELIMINARY; PRT; 506 AA.
AC Q9GR90;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE Putative mannose-specific lectin.
OS Polyandrocampa misakiensis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Styelidae; Polyandrocampa.
OX NCBI_TaxID=7723;
RN [1]
RP SEQUENCE FROM N.A.
RA Sagara Y., Fujiwara S., Yubisui T.;
RL "cDNA clonings from Polyandrocampa misakiensis.";
DR EMBL; AB053114; BAB20045.1; -
DR InterPro; IPR005052; Lectin_leg.
DR Pfam; PF03388; Lectin_leg-like; 1.
KW Lectin.
SQ SEQUENCE 506 AA; 57282 MW; 81524C9295B9B199 CRC64;

Query Match 47.8%; Score 44; DB 5; Length 506;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 HLLQDDVSFPW 15

Db 40 HLIQEDLTVPFW 51

RESULT 10

O81301
ID O81301 PRELIMINARY; PRT; 518 AA.
AC O81301;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE T14P8.1 protein (Pectinesterase-like protein).
GN T14P8.1 OR AF4G02320.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA WASHU;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Kalicki J., Elliott G., Cloud J.;
RP "The sequence of A. thaliana T14P8.";
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF069298; AAC19272.1; -

DR EMBL; AL161494; CAB80725.1; -
DR InterPro; IPR000070; Pectinesterase.
DR Pfam; PF01095; Pectinesterase; 1.
DR PROSITE; PS00503; PECTINESTERASE_2; 1.
SQ SEQUENCE 518 AA; 57894 MW; BECBA58404C701B7 CRC64;

Query Match 47.3%; Score 43.5; DB 10; Length 518;
Best Local Similarity 38.9%; Pred. No. 42;
Matches 7; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 MPAHL---LQDDVSFPW 15
Db 171 IFGHIPGVKEDVGFPW 188

RESULT 11

O49796
ID O49796 PRELIMINARY; PRT; 136 AA.
AC O49796;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE B2126.F1_18.
GN MLCB2533.I3C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Hamlin N., Churcher C.M.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae.";
RL Mo., Microbiol. 7:197-206(1993).
DR EMBL; U00017; AAA17210.1; -
DR EMBL; AL035310; CAA22927.1; -
KW Hypothetical protein.
SQ SEQUENCE 136 AA; 14872 MW; F9DC6B023643ED34 CRC64;

Query Match 46.7%; Score 43; DB 2; Length 136;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAHLQDDVS 11
Db 36 PAHVLDQDLT 45

RESULT 12

O9H718
ID O9H718 PRELIMINARY; PRT; 208 AA.
AC O9H718;


```

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FLJ00096 protein (Fragment).
GN FLJ00096
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Chara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024492; BAB15782.1;
FT NON_TER
SQ SEQUENCE 208 AA; 22275 MW; 05781A54D80692C1 CRC64;

Query Match 46.7%; Score 43; DB 4; Length 208;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 PAHLQDDVSFPWP 16
Db 187 PSRLPDASNPWP 201

RESULT 13
Q99261 PRELIMINARY; PRT; 437 AA.
ID Q99261
AC Q99261
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 20, Last annotation update)
DE Putative chloride channel protein.
GN SPV1379.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Setate S., Suarov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006575; AAK34204.1;
DR InterPro; IPR001807; Cl-channel_volt.
DR Pfam; PF00654; voltage_CLC; 1.
DR PRINTS; PR00762; CLCHANNEL.
KW Complete proteome.
SQ SEQUENCE 437 AA; 47717 MW; 28DA453AC7EBC410 CRC64;

Query Match 46.7%; Score 43; DB 16; Length 437;
Best Local Similarity 54.5%; Pred. No. 43;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLOQDDVSFPWP 15
Db 117 ILOQDDVSFPW 127

RESULT 14
Q920X2 PRELIMINARY; PRT; 674 AA.
ID Q920X2
AC Q920X2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)

```

```

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Probable exinuclease ABC subunit C protein.
GN UVRC OR R01171 OR SMC00602.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [-]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Anpe F., Batut J.,
RA Boistard P., Becker A., Bounry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591786; SCA45750.1;
DR InterPro; IPR000445; Hhh.
DR InterPro; IPR003583; Hhh.1.
DR InterPro; IPR001943; UvrB/C.
DR InterPro; IPR004791; UvrC.
DR InterPro; IPR001162; UvrC.
DR InterPro; IPR000305; UvrC_N.
DR Pfam; PF01541; Exci_endo_N; 1.
DR Pfam; PF02151; UVR; 1.
DR ProDom; PD005870; UvrC_C; 1.
DR SMART; SM00278; Hhh1; 2.
DR TIGFAMS; TIGR00194; uvrC; 1.
KW Complete proteome.
SQ SEQUENCE 674 AA; 74417 MW; 14FCFE72FB39A580 CRC64;

Query Match 46.7%; Score 43; DB 16; Length 674;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 DDVSFPWP 16
Db 516 EDAGFPWP 524

RESULT 15
Q8S864 PRELIMINARY; PRT; 753 AA.
ID Q8S864
AC Q8S864;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 85.8 kDa protein.
GN OSJNBA0061K21.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC Spiegel L.A., Nascimiento L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Huang E.N., See L.H., Shah R.S., O'Shaughnessy A.,
RA Rodriguez M.A., Shekher M., Kirchoff K.A., Baker J.P., Schutz K.,
RA Dedhia N.N., McCombie W.R.;
RT "Genomic Sequence For Oryza sativa, Niponbare Strain, Clone
RT OSJNBA0061K21 From Chromosome 10, Complete Sequence.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA McCombie W.R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.

```

RA McCombie W.R.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Palmer L.E., Yu M., Spiegel L.A., Nascimento L.U., de la Bastide M.,
 RA Vil D.M., Preston R.R., Huang E.N., See L.H., Shah R.S.,
 RA O'Shaughnessy A., Rodriguez M.A., Shekher M., Kirchoff K.A.,
 RA Baker J.P., Schutz K., Bedhia N.N., McCombie W.R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AC016780; AAM08782.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 753 AA; 85824 MW; 34DFBC05BE47820E CRC64;

 Query Match 46.7%; Score 43; DB 10; Length 753;
 Best Local Similarity 53.3%; Pred. No. 76;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

 QY 1 MPAHLIQDDVSFPAP 15
 :||| ||| | | | |
 Db 705 VPASLEVDVDTSLPRW 719

Search completed: May 14, 2003, 15:30:54
 Job time : 37 secs